SORT

Statistics and Operations Research Transactions

Volume 28
Number 2, July - December 2004

ISSN: 1696-2281

Generalitat de Catalunya
Institut d'Estadística de Catalunya

Aims

SORT (*Statistics and Operations Research Transactions*) – formerly *Qüestiió* - is an international journal launched in 2003, published twice-yearly by the Institut d'Estadística de Catalunya (Idescat), co-sponsored by the Universitat Politècnica de Catalunya (UPC), Universitat de Barcelona, Universitat Autònoma de Barcelona and Universitat de Girona and with the co-operation of the Spanish Region of the International Biometric Society. SORT promotes the publication of original articles of a methodological or applied nature on statistics, operations research, official statistics and biometrics.

The journal is described in the *Encyclopedia of Statistical Sciences*, and referenced in the *Current Index to Statistics*, the *Índice Español de Ciencia y Tecnología*, *Statistical Methods and Abstracts*, as well as MathSci of the American Mathematical Society (*Current Mathematical Publications and Mathematical Reviews*).

SORT represents the third series of the *Quaderns d'Estadística i Investigació Operativa* (*Qüestiió*), published by Idescat since 1992 until 2002, which in turn followed from the first series *Quaderns d'Estadística, Sistemes, Informàtica i Investigació Operativa* (1977-1992). The three series of *Qüestiió* have their origin in the *Cuadernos de Estadística Aplicada e Investigación Operativa*, published by the UPC till 1977.

Editor in Chief

Carles M. Cuadras (*Universitat de Barcelona*)

Executive Editors

Joan del Castillo (*Universitat Autònoma de Barcelona*)
Guadalupe Gómez (*Universitat Politècnica de Catalunya*)
F. Javier Heredia (*Universitat Politècnica de Catalunya*)
Enric Ripoll (*Institut d'Estadística de Catalunya*)

Production Editor

Michael Greenacre (*Universitat Pompeu Fabra*)

Editorial Advisory Committee

John Aitchison	*University of Glasgow, Dept. of Statistics*
Jesús R. Artalejo	*Universidad Complutense, Depto. de Estadística e Investigación Operativa*
Manuel Artís	*Universitat de Barcelona, Dept. d'Econometria, Estadística i Economia Espanyola*
Jaume Barceló	*Universitat Politècnica de Catalunya, Dept. d'Estadística i Investigació Operativa*
M Jesús Bayarri	*Universitat de València, Dept. d'Estadística i Investigació Operativa*
Eduard Bonet	*ESADE-Universitat Ramon Llull, Dept. de Mètodes Quantitatius*
Laureano F. Escudero	*Universidad Miguel Hernández, Centro de Investigación Operativa*
Francisco R. Fernández	*Universidad de Sevilla, Dpto. de Estadística e Investigación Operativa*
Josep Fortiana	*Universitat de Barcelona, Dept. d'Estadística*
Robert Garfinkel	*University of Connecticut, Dept. of Operations and Information Management*
Ubaldo Garcia Palomares	*Universidad Simón Bolivar, Depto. de Procesos y Sistemas*
Montserrat Guillén	*Universitat de Barcelona, Dept. d'Econometria, Estadística i Economia Espanyola*
Montserrat Herrador	*Instituto Nacional de Estadística*
Pierre Joly	*École Nationale de la Statistique et de l'Analyse de l'Information*
Wojtek J. Krzanowski	*University of Exeter, Dept. of Statistics and Operational Research*
Stephen Lagakos	*University of Harvard, Dept. of Biostatistics*
Ludovic Lebart	*Centre Nationale de la Recherche Scientifique*
Manuel Martí	*Universitat Politècnica de Catalunya, Dept. d'Estadística i Investigació Operativa*
Heinz Neudecker	*University of Amsterdam, Dept. of Econometrics*
David Nualart	*Universitat de Barcelona, Dept. d'Estadística*
Josep M. Oller	*Universitat de Barcelona, Dept. d'Estadística*
Albert Prat	*Universitat Politècnica de Catalunya, Dept. d'Estadística i Investigació Operativa*
Javier Prieto	*Universidad Carlos III de Madrid, Depto. de Estadística y Econometría*
Pere Puig	*Universitat Autònoma de Barcelona, Dept. de Matemàtiques*
C. Radhakrishna Rao	*Penn State University, Center for Multivariate Analysis*
Albert Satorra	*Universitat Pompeu Fabra, Dept. d'Economia i Empresa*
Albert Sorribas	*Universitat de Lleida, Dept. de Ciències Mèdiques Bàsiques*
Michael Stephens	*Simon Fraser University, Dept. of Statistics & Actuarial Science*
Santiago Thió	*Universitat de Girona, Dept. d'Informàtica i Matemàtica Aplicada*
Frederic Utzet	*Universitat Autònoma de Barcelona, Dept. de Matemàtiques*
Vladimir Zaiats	*Universitat de Vic, Dept. d'Informàtica i Matemàtica*

ISSN: 1696-2281
SORT 28 (2) July - December 111-230 (2004)

SORT

Statistics and Operations Research Transactions

Sponsoring institutions

Universitat Politècnica de Catalunya
Universitat de Barcelona
Universitat de Girona
Universitat Autònoma de Barcelona
Institut d'Estadística de Catalunya

Supporting institution

Spanish Region of the International Biometric Society

Generalitat de Catalunya
Institut d'Estadística de Catalunya

SORT

Volume 28 (2), July-December 2004

Formerly Qüestiió

Contents

Articles

PAPER ECOLÒGIC

SORT 28 (2) July-December 2004, 111-124

On Invariant Density Estimation for Ergodic Diffusion Processes

Yu. A. Kutoyants*

Université du Maine

Abstract

We present a review of several results concerning invariant density estimation by observations of ergodic diffusion process and some related problems. In every problem we propose a lower minimax bound on the risks of all estimators and then we construct an asymptotically efficient estimator.

MSC: 62M05, 62G07, 62G20

Keywords: ergodic diffusion, density estimation, large sample theory, efficient estimation

1 Introduction

Suppose that we observe a trajectory $X^T = \{X_t, 0 \leq t \leq T\}$ of the diffusion process

$$\mathrm{d}X_t = S(X_t)\,\mathrm{d}t + \sigma(X_t)\,\mathrm{d}W_t, \quad X_0, \quad t \geq 0 \tag{1}$$

where the trend coefficient $S(\cdot)$ is an unknown function and the diffusion coefficient $\sigma(\cdot)^2$ is a known positive function. We assume that these functions are such that the equation (1) has a unique weak solution (see, e.g. Durret (1996). Moreover, we suppose that the following conditions are fulfilled.

This contribution has been presented in the Barcelona Conference on Asymptotic Statistics, which took place in Bellaterra (Barcelona, 2003).

Address for correspondence: Laboratoire de Statistique & Processus, Université du Maine, 72085 Le Mans, Cedex 9, FRANCE.

Received: November 2003
Accepted: June 2004

Conditions $\mathcal{A}_0$:

$$\int_0^x \exp\left\{-2\int_0^y \frac{S(v)}{\sigma(v)^2}\,\mathrm{d}v\right\}\mathrm{d}y \to \pm\infty, \quad \text{as} \quad x \to \pm\infty$$

and

$$G(S) = \int_{-\infty}^{\infty} \frac{1}{\sigma(x)^2}\exp\left\{2\int_0^x \frac{S(v)}{\sigma(v)^2}\,\mathrm{d}v\right\}\,\mathrm{d}x < \infty.$$

By these conditions the solution of equation (1) has ergodic properties, with the invariant density

$$f_S(x) = \frac{1}{G(S)\sigma(x)^2}\exp\left\{2\int_0^x \frac{S(v)}{\sigma(v)^2}\,\mathrm{d}v\right\}, \tag{2}$$

i.e., for any function $h(\cdot)$ such that $\mathbf{E}_S\,|h(\xi)| < \infty$ the *law of large numbers*

$$\frac{1}{T}\int_0^T h(X_t)\,\mathrm{d}t \longrightarrow \mathbf{E}_S\,h(\xi) \qquad \text{p.s.}$$

holds. We denote by ξ a random variable with density function $f_S(\cdot)$.

We consider the problem of estimation of the invariant density by observations X^T and study the properties of its estimators in the asymptotic of *large samples* $T \to \infty$. The initial value X_0 is supposed to have the same density function $f_S(\cdot)$, so the observed process is stationary and the observed value X_t is a random variable with density $f_S(\cdot)$. We recall that in i.i.d. case the density of one observation entirely defines the distribution of the whole sample, but in the case of continuous time stochastic processes the distribution of the observed trajectory is defined by all finite dimensional distributions. Hence the density of one observation does not identify the model. For an ergodic diffusion process this density nevertheless identifies the whole model, because the model is entirely defined by the trend (unknown) and diffusion (known) coefficients and having invariant density we can write the trend coefficient as

$$S(x) = \frac{\left(\sigma(x)^2 f_S(x)\right)'}{2f_S(x)}. \tag{3}$$

The problem of invariant density estimation was considered by many authors (see, e.g., Nguen (1979), Delecroix (1980), Castellana and Leadbetter (1986), Bosq (1998), van Zanten (2001) *et al.*). In particular, Castellana and Leadbetter (1986) showed that for any stationary process with one and two dimensional densities $f(y), f(\tau, y, z)$ under condition :

$C\mathcal{L}$. The functions $f(y), f(\tau, y, z), \tau > 0$ are continuous at point x and

$$|f(\tau, y, z) - f(y)f(z)| \le \psi(\tau) \in \mathcal{L}_1(\mathbf{R}_+), \tag{4}$$

this density $f(y)$ can be estimated with the *parametric rate* $\sqrt{T}$, i.e., let $\hat{f}_T(x)$ be a kernel type estimator

$$\hat{f}_T(x) = \frac{1}{T\varphi_T}\int_0^T K\left(\frac{X_t - x}{\varphi_T}\right) \mathrm{d}t \tag{5}$$

where $\varphi_T \to 0, T\varphi_T \to \infty$ and the kernel $K(\cdot)$ satisfies the usual properties, then

$$\lim_{T\to\infty} T\,\mathbf{E}\left(\hat{f}_T(x) - f(x)\right)^2 = A(x).$$

Here

$$A(x) = 2\int_0^\infty \left[f(\tau, x, x) - f(x)^2\right] \mathrm{d}\tau.$$

The condition $C\mathcal{L}$ can be verified for ergodic diffusion processes (see Veretennikov (1999) for sufficient conditions). Note that this verification requires much more regularity from the coefficients $S(\cdot)$ and $\sigma(\cdot)$, than we really need in this estimation problem.

In this work we propose several asymptotically efficient estimators of the density function without supposing that the condition $C\mathcal{L}$ is fulfilled.

2 Lower bound

We start with the minimax lower bound on the risks of all estimators. This bound was established for a wide class of loss functions (see Kutoyants (1997b), (1998)) but for simplicity of exposition we consider quadratic loss functions only.

Fix some $S_*(\cdot)$ and and $\delta > 0$ and introduce the set

$$V_\delta = \{S(\cdot)\ :\ \sup_{x\in\mathbf{R}} |S(x) - S_*(x)| \leq \delta\}.$$

The role of Fisher information in our problem plays the quantity

$$\mathrm{I}_f(S, x) = \left\{4\, f_S(x)^2\, \mathbf{E}_S\left(\frac{\chi_{\{\xi > x\}} - F_S(\xi)}{\sigma(\xi) f_S(\xi)}\right)^2\right\}^{-1}$$

where $F_S(\cdot)$ is the distribution function of the invariant law. Hence $F_S(\xi)$ is a uniform $[0, 1]$ random variable. We have the following result.

Theorem 1 *Let* $\sup_{S\in V_\delta} G(S) < \infty$, *and* $\mathrm{I}_f(S_*, x) > 0$, *then for all estimators* $\bar{f}_T(x)$

$$\underline{\lim}_{\delta\to 0}\ \underline{\lim}_{T\to\infty}\ \sup_{S(\cdot)\in V_\delta} T\,\mathbf{E}_S\left(\bar{f}_T(x) - f_S(x)\right)^2 \geq \mathrm{I}_f(S_*, x)^{-1}.$$

As usual in this type of problem (see Ibragimov and Khasminskii (1981), Chapter 4) the proof is based on the estimate

$$\sup_{S(\cdot)\in V_\delta} \mathbf{E}_S \left(\bar{f}_T(x) - f_S(x)\right)^2 \geq \sup_{\vartheta\in\Theta_\delta} \mathbf{E}_\vartheta \left(\bar{f}_T(x) - f_\vartheta(x)\right)^2$$

where the parametric sub-model corresponds to the trend coefficient $S(\vartheta, x) = S_*(x) + (\vartheta - \vartheta_0)\, \psi(x)\, \sigma(x)^2 \in V_\delta$, with the function $\psi(\cdot)$ from the class

$$\mathcal{K} = \left\{\psi(\cdot):\ \mathbf{E}_{S_*} \int_\xi^x \psi(v)\,\mathrm{d}v = (2\, f_{S_*}(x))^{-1}\right\}.$$

For this parametric family with Fisher information I_ψ we obtain a Hajek-Le Cam minimax bound and then choose the *least favourable family* (with minimal Fisher information) as follows

$$\inf_{\psi\in\mathcal{K}} \mathrm{I}_\psi = \mathrm{I}_f(S_*, x).$$

Note that for $\psi(\cdot) \in \mathcal{K}$ we have $f_\vartheta(x) = \vartheta + o(1)$ as $\delta \to 0$. The details of the proof can be found in Kutoyants (1997d), (1998), (2003).

This lower bound allows us define the asymptotically efficient estimator ϑ_T^* by the following equality:

$$\lim_{\delta\to 0} \lim_{T\to\infty} \sup_{S(\cdot)\in V_\delta} T\, \mathbf{E}_S \left(f_T^*(x) - f_S(x)\right)^2 = \mathrm{I}_f(S_*, x)^{-1}.$$

3 Asymptotically efficient estimators

We consider below three type of estimators: *local time, unbiased* and *kernel type.*

Local time estimator

Recall that local time $\Lambda_T(x)$ of the diffusion process (1) is defined by the formula

$$\Lambda_T(x) = \lim_{\varepsilon\downarrow 0} \frac{\mathrm{meas}\{t:\ |X_t - x| \leq \varepsilon,\ 0 \leq t \leq T\}}{4\,\varepsilon}$$

and it admits the representation (see Karatzas and Shreve (1991))

$$2\Lambda_T(x) = |X_T - x| - |X_0 - x| + \int_0^T \mathrm{sgn}\,(x - X_t)\ \mathrm{d}X_t$$

Local time estimator of the density is defined by the equality

$$f_T^\circ(x) = \frac{2\,\Lambda_T(x)}{\sigma(x)^2\,T}.$$

We study its asymptotic behavior under the following conditions:

$\mathcal{U}$. *The law of large numbers*

$$\mathbf{P}_S - \lim_{T\to\infty} \frac{4f_S(x)^2}{T} \int_0^T \left(\frac{\chi_{\{X_t>x\}} - F_S(X_t)}{\sigma(X_t)\,f_S(X_t)}\right)^2 \mathrm{d}t = \mathrm{I}_f(S,x)^{-1}$$

is uniform on $S(\cdot) \in V_\delta$.

$\mathcal{B}$.

$$\sup_{S(\cdot)\in V_\delta}\left(\mathbf{E}_S\left|\frac{\chi_{\{\xi>x\}} - F_S(\xi)}{\sigma(\xi)\,f_S(\xi)}\right|^2 + \mathbf{E}_S\left|\int_0^\xi \frac{\chi_{\{v>x\}} - F_S(v)}{\sigma(v)^2 f_S(v)}\mathrm{d}v\right|^2\right) < \infty.$$

Theorem 2 *Let the conditions $\mathcal{U},\mathcal{B}$ be fulfilled and $\mathrm{I}_f(S,x)$ be continuous on V_δ, then the estimator $f_T^\circ(x)$ is unbiased: $\mathbf{E}_S f_T^\circ(x) = f_S(x)$, asymptotically normal:*

$$\mathcal{L}_S\left\{\sqrt{T}\left(f_T^\circ(x) - f_S(x)\right)\right\} \Rightarrow \mathcal{N}\left(0, \mathrm{I}_f(S,x)^{-1}\right)$$

and is asymptotically efficient.

The proof follows directly from the representation

$$\begin{aligned}\sqrt{T}\left(f_T^\circ(x) - f_S(x)\right) &= \frac{2f(x)}{\sqrt{T}}\int_{X_0}^{X_T} \frac{\chi_{\{v>x\}} - F(v)}{\sigma(v)^2 f(v)}\,\mathrm{d}v \\ &\quad - \frac{2f(x)}{\sqrt{T}}\int_0^T \frac{\chi_{\{X_t>x\}} - F(X_t)}{\sigma(X_t)\,f(X_t)}\,\mathrm{d}W_t \end{aligned} \tag{6}$$

and the central limit theorem for stochastic integrals.

Unbiased estimator

Let us introduce the estimator of the density function

$$f_T^*(x) = \frac{1}{T}\int_0^T R_x(X_t)\,\mathrm{d}X_t + \frac{1}{T}\int_0^T N_x(X_t)\,\mathrm{d}t$$

where $h(\cdot) \in C'(\mathbf{R})$ and

$$R_x(y) = \frac{2\chi_{\{y<x\}} h(y)}{\sigma(x)^2\, h(x)}, \qquad N_x(y) = \frac{\chi_{\{y<x\}} h'(y)\sigma(y)^2}{\sigma(x)^2\, h(x)}.$$

Then it is easy to see that $\mathbf{E}_S f_T^*(x) = f_S(x)$ and

$$\mathcal{L}_S\left\{\sqrt{T}\left(f_T^*(x) - f_S(x)\right)\right\} \Rightarrow \mathcal{N}\left(0, \mathrm{I}_f(S,x)^{-1}\right).$$

The detailed proof (with conditions like $\mathcal{U}, \mathcal{B}$) can be found in Kutoyants (1998), (2003). It is based on the representation

$$\sqrt{T}\left(f_T^*(x) - f_S(x)\right) = \sqrt{T}\left(f_T^\circ(x) - f_S(x)\right) + o(1).$$

In particular, if $\sigma(x) \equiv 1$ and $h(x) = x^3$, then for $x \neq 0$

$$f_T^*(x) = \frac{2}{Tx^3}\int_0^T \chi_{\{X_t<x\}}\, X_t^3\, \mathrm{d}X_t + \frac{3}{Tx^3}\int_0^T \chi_{\{X_t<x\}}\, X_t^2\, \mathrm{d}t$$

is unbiased and asymptotically efficient estimator of the density.

Kernel type estimator

Let us introduce the *kernel type estimator*

$$\hat{f}_T(x) = \frac{1}{\sqrt{T}}\int_0^T K\left(\sqrt{T}(X_t - x)\right)\mathrm{d}t$$

where the kernel $K(\cdot)$ is a bounded function with compact support $[A, B]$ and

$$\int_A^B K(u)\,\mathrm{d}u = 1, \qquad \int_A^B u\,K(u)\,\mathrm{d}u = 0.$$

To study this estimator we need to suppose that the function $f_S(x)$ is continuously differentiable (the function $S(x)$ is continuous and $\sigma(x)^2$ is continuously differentiable). Then we obtain the representation

$$\sqrt{T}\left(\hat{f}_T(x) - f(x)\right) = \sqrt{T}\left(f_T^\circ(x) - f(x)\right) + o(1)$$

and the asymptotic normality of this estimator

$$\mathcal{L}_S\left\{\sqrt{T}\left(\hat{f}_T(x) - f_S(x)\right)\right\} \Rightarrow \mathcal{N}\left(0, \mathrm{I}_f(S,x)^{-1}\right)$$

follows from the asymptotic normality of the local time estimator. The detailed proof with exact conditions can be found in Kutoyants (1998), (2003). Using similar arguments we obtain its asymptotic efficiency. It is clear that $A(x) = \mathrm{I}_f(S,x)^{-1}$. The consistent estimation of the quantity $\mathrm{I}_f(S,x)^{-1}$ is proposed in Dehay and Kutoyants (2004).

4 Semiparametric estimation

Let us consider the problem of parameter

$$\vartheta_S = \mathbf{E}_S R(\xi) S(\xi) + \mathbf{E}_S N(\xi)$$

estimation by observations (1) (with unknown $S(\cdot)$ and known $\sigma(\cdot)^2$). Here $R(\cdot)$ and $N(\cdot)$ are known functions. We will see later that the problem of invariant density estimation is a particular case of this problem.

Introduce the Fisher information

$$\mathrm{I}_\vartheta(S) = \left\{\mathbf{E}_S\left(\frac{R(\xi)\sigma(\xi)^2 f_S(\xi) + 2M_S(\xi)}{\sigma(\xi) f_S(\xi)}\right)^2\right\}^{-1}$$

where

$$M_S(y) = \mathbf{E}\left(\left[F_S(y) - \chi_{\{\xi<y\}}\right][R(\xi)S(\xi) + N(\xi)]\right).$$

The first result is the lower bound similar to that of Theorem 1.

Theorem 3 *Let* $\sup_{S\in V_\delta} G(S) < \infty$ *and* $\mathrm{I}_\vartheta(S_*) > 0$, *then for all estimators* $\bar\vartheta_T$

$$\underline{\lim}_{\delta\to 0}\,\underline{\lim}_{T\to\infty}\,\sup_{S(\cdot)\in V_\delta} T\,\mathbf{E}_S\left(\bar\vartheta_T - \vartheta_S\right)^2 \geq \mathrm{I}_\vartheta(S_*)^{-1}.$$

The proof can be found in Kutoyants (1997c) and (2003).

By direct calculation we verify that the *empirical estimator*

$$\tilde\vartheta_T = \frac{1}{T}\int_0^T R(X_t)\,\mathrm{d}X_t + \frac{1}{T}\int_0^T N(X_t)\,\mathrm{d}t$$

(under moments conditions) is asymptotically normal:

$$\mathcal{L}_S\left\{\sqrt{T}\left(\tilde\vartheta_T - \vartheta_S\right)\right\} \Rightarrow \mathcal{N}\left(0, \mathrm{I}_\vartheta(S)^{-1}\right).$$

and asymptotically efficient:

$$\lim_{\delta\to 0}\lim_{T\to\infty}\sup_{S(\cdot)\in V_\delta} T\,\mathbf{E}_S\left(\tilde\vartheta_T-\vartheta_S\right)^2=\mathrm{I}_\vartheta(S_*)^{-1}.$$

Let us consider three different choices of the functions $R(\cdot)$ and $N(\cdot)$.

- **Distribution function** estimation. Let us put $R(y) = 0$ and $N(y) = \chi_{\{y<x\}}$, then $\vartheta_S = F_S(x)$. Hence the empirical distribution function

$$\tilde\vartheta_T=\hat F_T(x)=\frac{1}{T}\int_0^T\chi_{\{X_t<x\}}\,\mathrm{d}t$$

  is asymptotically efficient estimator of the invariant distribution function Kutoyants (1997a).

- **Density** estimation. Let us put $R(y) = \dfrac{\operatorname{sgn}(x-y)}{\sigma(x)^2}$ and $N(y) = 0$, then $\vartheta_S = f_S(x)$ and

$$\tilde\vartheta_T=\bar f_T(x)=\frac{1}{T\sigma(x)^2}\int_0^T\operatorname{sgn}(x-X_t)\,\mathrm{d}X_t.$$

  Therefore, $\bar f_T(x)$ is an asymptotically efficient estimator of the density.

- **Moments** estimation. Let us put $R(y) = 0$ and $N(y) = y^k$, then $\vartheta_S = \mathbf{E}_S\xi^k$ and the empirical moment

$$\tilde\vartheta_T=\frac{1}{T}\int_0^T X_t^k\,\mathrm{d}t$$

  is asymptotically efficient estimator of the moments of ergodic diffusion process.

5 Integral type risk

Let us consider integral type quadratic risk

$$\mathcal{R}\left(\bar f_T,f_S\right)=\mathbf{E}_S\int(\bar f_T(x)-f_S(x))^2\,\mathrm{d}x$$

and denote by

$$\mathcal{R}_f(S)=\int\mathrm{I}_f(S_*,x)^{-1}\,\mathrm{d}x$$

the limit value of this risk for local time estimator.

Introduce the condition

$$\sup_{S(\cdot)\in V_\delta}\left\{\int \mathbf{E}_S\left|\frac{\chi_{\{\xi>x\}}-F_S(\xi)}{\sigma(\xi)\,f_S(\xi)}\right|^2 \mathrm{d}x+\right.$$
$$\left.+\int \mathbf{E}_S\left|\int_0^\xi \frac{\chi_{\{v>x\}}-F_S(v)}{\sigma(v)^2 f_S(v)}\mathrm{d}v\right|^2 \mathrm{d}x\right\}<\infty. \tag{7}$$

Theorem 4 *Let the condition* (7) *be fulfilled, then*

$$\lim_{\delta\to 0}\lim_{T\to\infty}\inf_{\bar f_T}\sup_{S(\cdot)\in V_\delta} T\,\mathcal{R}\left(\bar f_T, f_S\right)=\mathcal{R}_f(S_*).$$

Note that this theorem contains two results. The first one is the lower bound for all estimators

$$\underline{\lim}_{\delta\to 0}\,\underline{\lim}_{T\to\infty}\sup_{S(\cdot)\in V_\delta} T\,\mathcal{R}\left(\bar f_T, f_S\right)\geq \mathcal{R}_f(S_*). \tag{8}$$

The upper bound we obtain with the help of local time estimator. Slight modification of the conditions allows obtain the same limit for the risks of unbiased and kernel type estimators. Therefore, all these estimators are asymptotically efficient in the sense of the bound (8). The proofs can be found in Kutoyants (2003). Note that Negri (2001) establishes the asymptotic efficiency of the local time estimator for the loss function with uniform metric, i.e., for $\mathbf{E}_S\ell\left(\sup_x \sqrt{T}\left|\bar f_T(x)-f_S(x)\right|\right)$.

6 Second order efficiency

Having so many asymptotically efficient estimators we seek now the second order efficient one. Let us study the quantity $\left(T\mathcal{R}\left(\bar f_T, f_S\right)-\mathcal{R}_f(S)\right)$. Note that for LTE

$$T^{\frac{1}{2}}\left(T\mathcal{R}\left(f_T^{\circ}, f_S\right)-\mathcal{R}_f(S)\right)\to\ Q\neq 0.$$

It can be shown that if the function $S(\cdot)$ is $k-1$ times differentiable then for certain kernel type estimators $\hat f_T(\cdot)$

$$T^{\frac{1}{2k-1}}\left(T\mathcal{R}\left(\hat f_T, f_S\right)-\mathcal{R}_f(S)\right)\to\ -P<0.$$

To answer the question why the rate $T^{\frac{1}{2k-1}}$ is better than $T^{\frac{1}{2}}$ we write the last expression as

$$T\mathcal{R}\left(\hat f_T, f_S\right)=\mathcal{R}_f(S)-P\,T^{-\frac{1}{2k-1}}(1+o(1))$$

and it is clear now that the slower rate is better. Now we can compare the different estimators by their constants P and the estimator with biggest P will be the best (second order efficient). Therefore, as before we have two problems: the first one is to obtain a lower bound on the risks of all estimators (to find the biggest P) and the second is to construct an estimator which attains this bound. The similar problem of second order efficient estimation was consider by Golubev and Levit (1996) and our proof follows the main steps of their work. Note that this result is in the spirit of Pinsker's approach (see Pinsker (1980)).

For simplicity of exposition we put $\sigma(\cdot) = 1$.

Theorem 5 (Dalalyan, Kutoyants (2003)) *Suppose that the function $S(\cdot)$ is $(k-1)$-times differentiable ($k > 1$), satisfies the condition*

$$\lim_{|x|\to\infty} \operatorname{sgn}(x)\, S(x) < 0,$$

where $S_*(x) = -x$ and belongs to the set

$$\Sigma_k = \left\{ S(\cdot) : \int \left[f_S^{(k)}(x) - f_{S_*}^{(k)}(x) \right]^2 \mathrm{d}x \le R \right\}.$$

Then for all estimators $\bar{f}_T(\cdot)$

$$\varliminf_{T\to\infty} \sup_{S(\cdot)\in\Sigma_k} T^{\frac{1}{2k-1}} \left[T\mathcal{R}\left(\bar{f}_T, f_S\right) - \mathcal{R}_f(S) \right] \ge -\hat{\Pi}(k,R)$$

where

$$\hat{\Pi}(k,R) = 2(2k-1)\left(\frac{4k}{\pi(k-1)(2k-1)}\right)^{\frac{2k}{2k-1}} R^{-\frac{1}{2k-1}}.$$

The proof can also be found in Kutoyants (2003).

Let us introduce a subdivision of $\mathbf{R}$ on intervals $I_m = [a_m - \delta_T, a_m + \delta_T]$, where $a_m = 2m\delta_T, m = 0, \pm 1, \pm 2, \ldots$ and $\delta_T \to 0$. The asymptotically second order efficient estimator (for $x \in I_m$) can be written as

$$\hat{f}_T(x) = \frac{1}{2T\delta_T} \int_0^T \sum_{l=-\hat{\nu}_T}^{\hat{\nu}_T} \left(1 - \left|\frac{l}{\hat{\nu}_T}\right|^{k_T}\right) \cos\left(\frac{\pi l(x - X_t)}{\delta_T}\right) \chi_{\{X_t\in I_m\}} \mathrm{d}t$$

or

$$\hat{f}_T(x) = \sum_{l=-\hat{\nu}_T}^{\hat{\nu}_T} \left(1 - \left|\frac{l}{\hat{\nu}_T}\right|^{k_T}\right) \frac{1}{2\delta_T} \int_{a_m-\delta_T}^{a_m+\delta_T} \cos\left(\frac{\pi l(x-y)}{\delta_T}\right) f_T^\circ(y)\, \mathrm{d}y$$

Here $k_T = k + \mu_T, \mu_T = 1/\sqrt{\log T} \to 0$,

$$\hat{v}_T = \delta_T \left(\frac{8k\pi^{2(k-1)}}{RT(k-1)(2k-1)} \right)^{-\frac{1}{2k-1}} \to \infty$$

Theorem 6 (Dalalyan, Kutoyants (2004)) *Let the conditions of Theorem 5 be fulfilled, then*

$$\lim_{T\to\infty} \sup_{S(\cdot)\in\Sigma_k} T^{\frac{1}{2k-1}} \left[T\mathcal{R}\left(\hat{f}_T, f_S\right) - \mathcal{R}_f(S) \right] = -\hat{\Pi}(k, R).$$

The proof can also be found in Kutoyants (2003).

7 Trend estimation

Let us consider the problem of trend coefficient estimation. As before the observed process (1) is ergodic diffusion with unknown $S(\cdot)$ and known diffusion coefficient, which we put (for simplicity of exposition) to be equal 1. The problem of trend estimation was studied by several authors (see, e.g., Banon (1978), Galtchouk and Pergamenshchikov (2001)). Therefore we observe a trajectory $X^T = \{X_t, 0 \le t \le T\}$ of the solution of the stochastic differential equation

$$\mathrm{d}X_t = S(X_t)\,\mathrm{d}t + \mathrm{d}W_t, \quad X_0, \quad 0 \le t \le T.$$

The trend coefficient can be written with the help of invariant density $f_S(\cdot)$ as

$$S(x) = \frac{f'_S(x)}{2\, f'_S(x)}. \tag{9}$$

Hence for for estimation of $S(x)$ we can use the estimators of density and its derivative. The error of the estimators we measure with the help of the following risk

$$\mathcal{R}\left(\bar{S}_T, S\right) = \mathbf{E}_S \int \left(\bar{S}_T(x) - S(x)\right)^2 f_S(x)^2\,\mathrm{d}x.$$

The conditions of the regularity are similar to that of the Section 5.

Conditions $\mathcal{S}_\delta$.

$\mathcal{S}_1$. *The function $S(\cdot)$ has polynomial majorant and*

$$\overline{\lim_{|x|\to\infty}} \operatorname{sgn}(x)\ S(x) < 0.$$

$\mathcal{S}_2$. *The function $S(\cdot) \in C^k(\mathbf{R})$ with some $k \geq 1$ and belongs to the set*

$$\Sigma_\delta = \left\{ S(\cdot) \in V_\delta \; : \; \int_{\mathbf{R}} \left[f_S^{(k+1)}(x) - f_{S_*}^{(k+1)}(x) \right]^2 \, \mathrm{d}x \leq 4\, R \right\}.$$

$\mathcal{S}_3$. *The Fourier transform $\varphi_*(\cdot)$ of the function $f'_{S_*}(\cdot)$ is such that*

$$\int_{\mathbf{R}} |\lambda|^{2k+\tau} \left| \varphi_*(\lambda) \right|^2 \mathrm{d}\lambda < \infty$$

with some positive constant τ.

Let us put

$$\Pi(k, R) = (2k+1) \left(\frac{k}{\pi (k+1)(2k+1)} \right)^{\frac{2k}{2k+1}} R^{\frac{1}{2k+1}}$$

The first result is the minimax lower bound.

Theorem 7 (Dalalyan, Kutoyants (2002)) *Let the conditions $\mathcal{S}_\delta$ be fulfilled. Then for any estimator $\bar{S}_T(\cdot)$ we have*

$$\varliminf_{T\to\infty} \inf_{\bar{S}_T} \sup_{S(\cdot)\in\Sigma_\delta} T^{\frac{2k}{2k+1}} \mathcal{R}\left(\bar{S}_T, S\right) \geq \Pi(k, R).$$

According to (9) we introduce the estimator

$$\hat{S}_T(x) = \frac{\hat{\vartheta}_T(x)}{2\, f_T^\circ(x) + \varepsilon_T\, e^{-l_T |x|}}$$

where $f_T^\circ(x)$ is the local-time estimator of the density, $\varepsilon_T = T^{-(1-\kappa)/2}$, $l_T = [\ln T]^{-1}$, the constant $\kappa < 1/(2k+1)$ and $\hat{\vartheta}_T(x)$ is the asymptotically efficient estimator of the derivative $f'_S(x)$ Dalalyan, Kutoyants (2003):

$$\hat{\vartheta}_T(x) = \frac{2\nu_T}{T} \int_0^T K^*\left(\nu_T(x - X_t)\right) \mathrm{d}X_t$$

where the kernel

$$K^*(x) = \frac{1}{\pi} \int_0^1 \left(1 - u^{k+\mu_T}\right) \cos(ux)\, \mathrm{d}u$$

and

$$\nu_T = \left(\frac{\pi\, R\, (k+1)(2k+1)}{4k} \right)^{\frac{1}{2k+1}} T^{\frac{1}{2k+1}}.$$

Here $\mu_T = (\log T)^{-1/2}$.

Theorem 8 (Dalalyan, Kutoyants (2002)) *Let the conditions $\mathcal{S}_\delta$ be fulfilled then*

$$\lim_{T\to\infty} \sup_{S(\cdot)\in\Sigma_\delta} T^{\frac{2k}{2k+1}} \mathcal{R}\left(\hat{S}_T, S\right) = \Pi(k,R).$$

If the values $k \geq 2$ and $R > 0$ are unknown then it is possible to construct an adaptive estimator $\tilde{S}_T(\cdot)$, which has the same asymptotic properties as $\hat{S}_T(\cdot)$.

Theorem 9 (Dalalyan (2003)) *Let the conditions $\mathcal{S}_\delta$ be fulfilled then*

$$\lim_{T\to\infty} \sup_{S(\cdot)\in\Sigma_\delta} T^{\frac{2k}{2k+1}} \mathcal{R}\left(\tilde{S}_T, S\right) = \Pi(k,R).$$

8 References

Banon, G. (1978). Non parametric identification for diffusion processes. *SIAM. J. Control Optim.*, 16, 3, 380-395.

Bosq, D., Davydov, Y. (1998). Local time and density estimation in continuous time. *Math. Methods Statist.*, 8, 1, 22-45.

Bosq, D. (1998). *Nonparametric Statistics for Stochastic Processes.* (Second edition), Lecture Notes Statist., 110, New York: Springer-Verlag.

Castellana, J. V., Leadbetter, M. R. (1986). On smoothed density estimation for stationary processes. *Stochastic Process. Appl.*, 21, 179-193.

Dalalyan, A. (2002). Sharp adaptive estimation of the trend coefficient for ergodic diffusion, preprint n.º 02-1, Laboratoire Statistique et Processus, Université du Maine, France.

Dalalyan, A., Kutoyants, Yu. A. (2002). Asymptotically efficient trend coefficient estimation for ergodic diffusion. *Math. Methods Statist.*, 11, 4, 402-427.

Dalalyan, A., Kutoyants, Yu. A. (2003). Asymptotically efficient estimation of the derivative of invariant density. *Stat. Inference Stoch. Process.*, 6, 1, 89-107.

Dalalyan, A., Kutoyants, Yu. A. (2004). On second order minimax estimation of invariant density for ergodic diffusion. *Statistics & Decisions*, 22, 17-41.

Dehay, D., Kutoyants, Yu. A. (2004). On confidence intervals for distribution function and density of ergodic diffusion process. *J. Statist. Plann. Inference*, 124, 1, 63-73.

Delecroix, M. (1980). Sur l'estimation des densités d'un processus stationnaire à temps continu. *Publications de l'ISUP*, XXV, 1-2, 17-39.

Durett, R. (1996). *Stochastic Calculus: A Practical Introduction.* Boca Raton: CRC Press.

Galtchouk, L. and Pergamenshchikov, S. (2001). Sequential nonparametric adaptive estimation of the drift coefficient in diffusion processes. *Math. Methods Statist.*, 10, 3, 316-330.

Golubev, G. K., Levit, B. Ya. (1996). On the second order minimax estimation of distribution functions. *Math. Methods Statist.*, 5, 1, 1-31.

Ibragimov, I. A., Khasminskii, R. Z. (1981). *Statistical Estimation. Asymptotic Theory.* New York: Springer-Verlag.

Karatzas, I., Shreve, S. E. (1991). *Brownian Motion and Stochastic Calculus.* New York: Springer-Verlag.

Kutoyants, Yu. A. (1997*a*). Efficiency of empiric distribution function for ergodic diffusion processes. *Bernoulli*, 3 (4), 445-456.

Kutoyants, Yu. A. (1997*b*). Some problems of nonparametric estimation by the observations of ergodic diffusion processes. *Statist. Probab. Lett.*, 32, 311-320.

Kutoyants Yu. A. (1997*c*). On semiparametric estimation for ergodic diffusion. *Proc. A. Razmadze Math. Inst. (in honor of R. Chitashvili)*, Tbilisi, 115, 45-58.

Kutoyants, Yu. A. (1997*d*). On density estimation by the observations of ergodic diffusion processes. In *Statistics and Control of Stochastic Processes*, The Liptser Festschrift, Yu. M. Kabanov, B. L. Rozovskii, A. N. Shiryaev (eds), Singapore: World Scientific, 253-274.

Kutoyants, Yu. A. (1998). Efficient density estimation for ergodic diffusion process. *Stat. Inference Stoch. Process.*, 1, 2, 131-155.

Kutoyants, Yu. A. (2003). *Statistical Inference for Ergodic Diffusion Processes*. New York: Springer.

Kutoyants, Yu. A., Negri, I. (2001). On L_2 efficiency of empiric distribution of ergodic diffusion processs. *Theory Probab. Appl.*, 46, 1, 164-169.

Negri, I. (2001). On efficient estimation of invariant density for ergodic diffusion processes. *Statist. Probab. Lett.*, 51, 1, 79-85.

Nguyen, H. T. (1979). Density estimation in a continuous-time Markov processes. *Ann. Statist.*, 7, 341-348.

Pinsker, M. S. (1980). Optimal filtering of square integrable signals in Gaussian white noise. *Probl. Inf. Transm.*, 16, 120-133.

Veretennikov, A. Yu. (1999). On Castellana-Leadbetter's condition for diffusion density estimation. *Stat. Inference Stoch. Process.*, 2, 1, 1-9.

van Zanten, H. (2001). Rates of convergence and asymptotic normality of kernel estimators for ergodic diffusion processes. *Nonparametric Statist.*, 13 (6), 833-850.

SORT 28 (2) July-December 2004, 125-160

Robust estimation and forecasting for beta-mixed hierarchical models of grouped binary data

M. A. Pashkevich, Yu. S. Kharin*

Belarusian State University

Abstract

The paper focuses on robust estimation and forecasting techniques for grouped binary data with misclassified responses. It is assumed that the data are described by the beta-mixed hierarchical model (the beta-binomial or the beta-logistic), while the misclassifications are caused by the stochastic additive distortions of binary observations. For these models, the effect of ignoring the misclassifications is evaluated and expressions for the biases of the method-of-moments estimators and maximum likelihood estimators, as well as expressions for the increase in the mean square error of forecasting for the Bayes predictor are given. To compensate the misclassification effects, new consistent estimators and a new Bayes predictor, which take into account the distortion model, are constructed. The robustness of the developed techniques is demonstrated via computer simulations and a real-life case study.

MSC: 62F35

Keywords: grouped binary data, distortions, hierarchical models, beta-binomial, beta-logistic, robust, estimation, forecasting

1 Introduction

Grouped binary data frequently arise in longitudinal studies that are carried out over a group of similar objects (Diggle *et al.* 2002). A natural way to describe this kind of data is using the binomial model (Collet 2002). However, the binomial model often leads to inaccurate statistical inference due to the so called "over-dispersion" effects (Brooks 2001). These effects may occur for two main reasons (Neuhaus 2002): (i) intergroup correlation, i.e. violation of the independence assumption of the experiment outcomes

**Address for correspondence:* Department of Mathematical Modeling and Data Analysis. Faculty of Applied Mathematics and Computer Science. Belarusian State University. 4 F.Skaryny Ave., Minsk 220050, Belarus. pma@omegasoftware.com, kharin@bsu.by

Received: October 2003
Accepted: June 2004

for a particular object, and (ii) intragroup correlation caused by the heterogeneity among objects. So, special "random effects" models are used to describe the heterogeneity and correlated outcomes (Coull and Agresti 2000).

The beta-mixed hierarchical models of grouped binary data are widely used in practical applications when information about experiment conditions is not available. The most popular models of this class are the beta-binomial model (BBM) that supposes that the data on object properties are not available, and the beta-logistic model (BLM) that supposes that they are known. The BBM was originally proposed by Pearson (1925), formalized by Skellam (1948) and is associated with many useful results in applied statistics due to its conjugate property (Prentice 1988) that allows avoiding numerical integration while using Bayes approach for forecasting of response probabilities (Slaton *et al.* 2000). The BLM is an extension of the BBM that was proposed by Heckman and Willis (1977); it is widely used in economics, biometrics, political sciences and other applications (Pfeifer 1998; Nathan 1999).

In real life, the observed binary outcomes are often misclassified (Neuhaus 1999), and the classical statistical procedures that are optimal for the hypothetical model may lose their "good" properties under distortions (Kharin 1996). Hence, it is important to analyze the sensitivity of the classical estimators and predictors w.r.t. response misclassifications and, if needed, to develop new statistical procedures that are robust to these distortions (Huber 1981; Hampel *et al.* 1986). Although a number of papers have been published on robustness of the linear mixed model (Gill 2001), logistic regression (Kordzakhia *et al.* 2001), binomial model (Ruckstuhl and Welsh 2001), inference for dichotomous survey data (Gaba and Winkler 1992), and on the Bayesian identifiability problem of multinimial data with misclassifications (Swarzt *et al.* 2004), these results can not be directly applied to the grouped binary data due to their specific property.

The literature review shows that little research has been done on investigation the robustness issue for the special models of the grouped binary data. The major contribution to this domain has been done by Neuhaus, who has extended his general results for the binary regression models under response misclassifications (Neuhaus 1999) to the clustered and longitudinal binary data case. In his recent work, Neuhaus (2002) obtained expressions for the parameter bias and developed methods for consistent estimation for the population-averaged models (Liang and Zeger 1986). He also examined a special case of the cluster-specific models (Zeger and Karim 1991), the logistic normal model, which is an extension of the logistic regression to the grouped binary data case. However, as noted by Neuhaus (2002), "the derivation of bias expressions for nonlogistic links will require a different approach than for the logistic" since the specific property of the logistic link function was used to obtain the expressions.

This paper focuses on the robustness issues for the beta-mixed hierarchical models under stochastic additive distortions of binary observations. These models belong to the cluster-specific type but have not been addressed in the related works yet. The remainder of the paper is organized as follows. Section 2 is devoted to the problem statement and definition of the related mathematical models. Section 3 concentrates on the robust estimation of the beta-binomial model parameters, while Section 4 deals with the same problem for the beta-logistic model. Section 5 is dedicated to the robust forecasting based on the beta-mixed hierarchical models (both beta-binomial and beta-logistic ones). Section 6 presents an application example and evaluation of the developed methods for a real-life case study. Finally, Section 7 summarizes the main contributions of the paper.

2 Mathematical models and research problems

Let us consider k clusters with the covariates $Z_i \in R^m, i = 1, \ldots, k$, and let $B_i = (B_{i1}, B_{i2}, \ldots, B_{in_i}) \in \{0, 1\}^{n_i}$ be the binary responses of n_i Bernoulli trials over the cluster i. Let us also assume that the following two assumptions hold.

$A1$. Within the cluster i, the success probability p_i is a random variable that follows the beta distribution with the true unknown parameters $\alpha_i^0 = f_\alpha(Z_i)$, $\beta_i^0 = f_\beta(Z_i)$, where $f_\alpha(.) : R^m \to R^+$, $f_\beta(.) : R^m \to R^+$.

$A2$. Random variables $p_1, p_2, \ldots, p_k$ are independent in total.

Let us refer to the defined above data model as the beta-mixed hierarchical model of the grouped binary data. In this paper, we focus on two models of this type that are frequently used in practical applications (the beta-binomial and the beta-logistic), which are specified as follows:

$$\begin{aligned} &\text{BBM:} \quad f_\alpha(Z_i) = \alpha^0, \;\; f_\beta(Z_i) = \beta^0, \;\; n_i = n; \\ &\qquad\quad \text{model parameters: } n \in N, \;\; \alpha^0, \beta^0 \in R\,. \end{aligned}$$

$$\begin{aligned} &\text{BLM:} \quad f_\alpha(Z_i) = \exp(Z_i^T a^0), \; f_\beta(Z_i) = \exp(Z_i^T b^0); \\ &\qquad\quad \text{model parameters: } n_1, \ldots, n_k \in N, \;\; a^0, b^0 \in R^m. \end{aligned}$$

For the BBM, it is assumed that the number of Bernoulli trials $n_i = n$ is the same for all clusters and n is known a priori. Estimation of the remaining BBM parameters α^0, β^0 is performed (Tripathi *et al.* 1994) using the method of moments (explicit expressions) or the method of maximum likelihood (numerical algorithm). For the BLM, the number of Bernoulli trials n_i may vary across the clusters and is also known a priori, while the other parameters a^0, b^0 are estimated using the maximum likelihood numerical algorithm (Slaton *et al.* 2000).

One of the main problems for the grouped binary data that is strongly motivated by practical applications, is the forecasting of the success probabilities $p_1, \ldots, p_k$ for for the future trials using the past binary outcomes $B = \{B_1, \ldots, B_k\}$ obtained for small sample sizes n_i that are too small to have accurate traditional estimator $\hat{p}_i = n_i^{-1} x_i^0$ (Collet 2002). For the beta-mixed hierarchical models, this problem is solved via the Bayes predictor function (Diggle *et al.* 2002)

$$\tilde{p}_i(x_i^0) = (\alpha_i^0 + x_i^0)/(\alpha_i^0 + \beta_i^0 + n_i), \tag{1}$$

where $x_i^0 = \sum_{j=1}^{n_i} B_{ij}$, $i = 1, 2, \dots, k$, are the sums of the binary outcomes within the cluster. This predictor ensures the minimal mean square error of forecasting when the consistent estimators of the model parameters $\{\alpha_i^0, \beta_i^0\}$ are used.

Suppose now that the original binary data B are contaminated by the stochastic additive binary distortions $\{\eta_{ij}\}$, and we observe the distorted binary responses $\tilde{B}$

$$\tilde{B}_{ij} = B_{ij} \oplus \eta_{ij} \tag{2}$$

with the misclassifications defined as

$$\mathbf{P}\{\tilde{B}_{ij} = 1 | B_{ij} = 0\} = \varepsilon_0, \;\; \mathbf{P}\{\tilde{B}_{ij} = 0 | B_{ij} = 1\} = \varepsilon_1, \tag{3}$$

where $\oplus$ is the modulo 2 sum, and $\varepsilon_0, \varepsilon_1 \ll 1$ are the distortion levels which can be either known or unknown (Copas 1988). In this settings, two main research problems arise:

(i) Evaluation of the effects of ignoring the misclassifications for the classical model parameter estimation techniques and response probability forecasting methods.

(ii) Construction of new estimation and prediction methods, which take into account the distortion model and compensate the misclassification effect.

In the remaining sections, these problems are solved separately for the BBM and BLM parameter estimation, while the forecasting is examined and enhanced simultaneously for both of them. For the first problem, the estimation bias and the increase in the mean square error of forecasting are evaluated via asymptotic expansions. For the second one, new estimation and forecasting methods, which are based on the obtained probability distribution of the distorted data, are proposed.

It should be noted that for the BBM (Section 3), the paper considers the case of equal group sizes since it is typical for many application areas that exploit this model. The assumption $n_i = n$ allows obtaining simple expressions and helps to develop intuition about the distortions influence on the BBM inference. However, the results for the BBM with different $\{n_i\}$ can be easily obtained as a special case of the BLM results (Section 4), where the covariates $\{Z_i\}$ are the same for all clusters.

For further convenience, let us introduce the following notation: MM-estimator – the method of moments estimator, ML-estimator– the method of maximum likelihood estimator, $o(\varepsilon)$, $O(\varepsilon)$ –Landau symbols for $\varepsilon \to 0$, $Y_n = O_P(Z_n)$– probability Landau symbol for random sequences $Y_n, Z_n \in R$. The detailed definition of $O_P(.)$ and the proofs of theorems are given in Mathematical Appendix.

3 Robust estimation of the beta-binomial model

Distorted beta-binomial distribution. Let x_i be the number of successes for the i-th cluster: $x_i = \sum_{j=1}^{n_i} \tilde{B}_{ij}, \quad i = 1, \dots, k$. The following theorem defines the probability distribution of the random variable x_i under the distortions (2), (3).

Theorem 1 *The probability distribution of the distorted beta-binomial random variable x_i can be represented as a weighted sum*

$$P_r(\alpha, \beta, \varepsilon_0, \varepsilon_1) = \sum_{s=0}^{n} w_{rs}(\varepsilon_0, \varepsilon_1) \cdot P_s^0(\alpha, \beta), \tag{4}$$

where $\{P_s^0\}$ are the non-distorted probabilities for the BBM with the parameters n, α, β

$$P_s^0(\alpha, \beta) = \binom{n}{s} \frac{B(\alpha + s, \beta + n - s)}{B(\alpha, \beta)},$$

$B(.)$ is the complete beta function, and the weights for the distortion levels $\varepsilon_0, \varepsilon_1$ are computed as

$$w_{rs}(\varepsilon_0, \varepsilon_1) = \sum_{l=\max(s,r)}^{\min(n,s+r)} \binom{s}{l-r} \binom{n-s}{l-s} \varepsilon_0^{l-s} (1-\varepsilon_0)^{n-l} \varepsilon_1^{l-r} (1-\varepsilon_1)^{s+r-l}, \quad s, r = 0, 1, \dots, n.$$

Using this theorem, it can be proved that the mean and variance of the distribution (4) are

$$\mathbf{E}\{x_i\} = \varepsilon_0 \frac{n\beta}{\alpha+\beta} + (1-\varepsilon_1) \frac{n\alpha}{\alpha+\beta}, \qquad \mathbf{V}\{x_i\} = \varepsilon_0(1-\varepsilon_0) \frac{n\beta}{\alpha+\beta} + \varepsilon_1(1-\varepsilon_1) \frac{n\alpha}{\alpha+\beta} + (1-\varepsilon_0-\varepsilon_1)^2 \cdot V_0,$$

where $V_0 = (n\alpha\beta(\alpha+\beta+n))/((\alpha+\beta)^2(\alpha+\beta+1))$ is the variance of the non-distorted BBM. Let us refer to the distribution (4) as the distorted beta-binomial distribution (DBBD) with the parameters n, α, β, ε_0, ε_1.

As follows from the theorem proof (see Appendix), the weights w_{rs} can be treated as the probabilities that the distorted value r was originated from the non-distorted sum of the binary outcomes s. It should be noted that when $\varepsilon_0 = \varepsilon_1 = 0$, the proposed distribution (4) is identical to the classical beta-binomial distribution (BBD) with the parameters n, α, β, and the weight matrix $W = (w_{rs})$ is the identity one. If the distortion levels are small, the matrix W can be approximated by the asymptotic expansion

$$W(\varepsilon_0, \varepsilon_1) = I + W'_{\varepsilon_0} \cdot \varepsilon_0 + W'_{\varepsilon_1} \cdot \varepsilon_1 + o(\varepsilon_0, \varepsilon_1), \tag{5}$$

where I is the identity matrix, and the matrices W'_{ε_0}, W'_{ε_1} are calculated as

$$W'_{\varepsilon_0} = \begin{pmatrix} -n & 0 & 0 & \dots & 0 \\ n & -(n-1) & 0 & \dots & 0 \\ 0 & n-1 & -(n-2) & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & 0 \end{pmatrix}, \quad W'_{\varepsilon_1} = \begin{pmatrix} 0 & 1 & 0 & \dots & 0 \\ 0 & -1 & 2 & \dots & 0 \\ 0 & 0 & -2 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & -n \end{pmatrix}.$$

The expression (5) allows obtaining the following asymptotic relation between the distorted $P_r(\varepsilon_0, \varepsilon_1)$ and the original P_r^0 probabilities

$$P_r(\varepsilon_0, \varepsilon_1) = P_r^0 + \left((n-r+1)P_{r-1}^0 - (n-r)P_r^0\right) \cdot \varepsilon_0 + \left((r+1)P_{r+1}^0 - rP_r^0\right) \cdot \varepsilon_1 + o(\varepsilon_0, \varepsilon_1), \tag{6}$$

where $P_{-1}^0 = P_{n+1}^0 = 0$. This expression can be employed to assess the sensitivity of the beta-binomial distribution to the distortions (2), (3). In the following subsection, the result of Theorem 1 and the expression (6) are used to evaluate the sensitivity of the classical BBM estimators.

Robustness of the classical estimators. Let α^0,β^0 be the true unknown values of the BBM parameters, and let $\Delta\tilde{\alpha}(\varepsilon_0, \varepsilon_1)$, $\Delta\tilde{\beta}(\varepsilon_0, \varepsilon_1)$ be the biases of the parameter estimators that ignore the misclassifications with the levels $\varepsilon_0, \varepsilon_1$. The following theorems evaluate the robustness of the classical MM and ML-estimators via their biases w.r.t. the distortion levels.

Theorem 2 *The bias of the classical MM-estimator of the BBM parameters, which ignores the misclassifications, satisfies the following asymptotic expansion*

$$\begin{pmatrix} \Delta\tilde{\alpha}_{MM} \\ \Delta\tilde{\beta}_{MM} \end{pmatrix} = \begin{pmatrix} \alpha^0 + 2\beta^0 + 1 & \alpha^0(\alpha^0+1)/\beta^0 \\ \beta^0(\beta^0+1)/\alpha^0 & 2\alpha^0 + \beta^0 + 1 \end{pmatrix} \cdot \begin{pmatrix} \varepsilon_0 \\ \varepsilon_1 \end{pmatrix} + \begin{pmatrix} o(\varepsilon_0, \varepsilon_1) + O_P(1/\sqrt{k}) \\ o(\varepsilon_0, \varepsilon_1) + O_P(1/\sqrt{k}) \end{pmatrix}. \tag{7}$$

Theorem 3 *The bias of the classical ML-estimator of the BBM parameters, which ignores the misclassifications, satisfies the asymptotic expansion*

$$\begin{pmatrix} \Delta\tilde{\alpha}_{ML} \\ \Delta\tilde{\beta}_{ML} \end{pmatrix} = \begin{pmatrix} H_{11} & H_{12} \\ H_{21} & H_{22} \end{pmatrix}^{-1} \begin{pmatrix} G_{11} & G_{12} \\ G_{21} & G_{22} \end{pmatrix} \cdot \begin{pmatrix} \varepsilon_0 \\ \varepsilon_1 \end{pmatrix} + \begin{pmatrix} o(\varepsilon_0, \varepsilon_1) + O_P(1/\sqrt{k}) \\ o(\varepsilon_0, \varepsilon_1) + O_P(1/\sqrt{k}) \end{pmatrix}, \tag{8}$$

where explicit expressions for the matrices H, G are given in Mathematical Appendix.

As follows from these theorems, the classical MM and ML-estimators of the BBM parameters become biased and inconsistent under the distortions. Expressions (7), (8) allow assessing the sensitivity of these estimators to the misclassifications (3). Let us now construct consistent and unbiased estimators that take into account the distortion model (2), (3).

Robust estimation in the case of known distortion levels. Let us consider the case when the distortion levels $\varepsilon_0, \varepsilon_1$ are known a priori. Denote the empirical moments of the order r as $m_r^* = k^{-1} \sum_{i=1}^k x_i^r$. The following theorems define consistent and asymptotically unbiased MM and ML-estimators for the case of known $\varepsilon_0, \varepsilon_1$.

Theorem 4 *The consistent and asymptotically unbiased MM-estimator, which takes into account the distortion model (2), (3), is expressed as*

$$\hat{\alpha}_{MM} = \frac{\delta_\alpha(m_1^*, \varepsilon_0) \cdot \mu(m_1^*, m_2^*, \varepsilon_0, \varepsilon_1)}{\Delta(m_1^*, m_2^*, \varepsilon_0, \varepsilon_1)}, \quad \hat{\beta}_{MM} = \frac{\delta_\beta(m_1^*, \varepsilon_1) \cdot \mu(m_1^*, m_2^*, \varepsilon_0, \varepsilon_1)}{\Delta(m_1^*, m_2^*, \varepsilon_0, \varepsilon_1)}, \tag{9}$$

where

$$\delta_\alpha = m_1^* - n\varepsilon_0, \quad \delta_\beta = n - m_1^* - n\varepsilon_1, \quad \mu = m_1^* n - m_2^* - (\varepsilon_0 \delta_\beta + m_1^* \varepsilon_1)(n-1),$$

$$\Delta = (1 - \varepsilon_1 - \varepsilon_0)\left(m_2^* n - m_1^* n - m_1^{*2}(n-1)\right).$$

Theorem 5 *The consistent and asymptotically unbiased ML-estimator, which takes into account the distortion model (2), (3), can be derived by applying the classical ML-estimator to the filtered empirical probabilities*

$$\hat{P}_r^0 = \sum_{l=0}^n v_{rl}(\varepsilon_0, \varepsilon_1) \cdot \hat{P}_l(\varepsilon_0, \varepsilon_1), \tag{10}$$

where $\{\hat{P}_0, \ldots, \hat{P}_n\}$ *is the empirical probability distribution of the distorted sample* $\{x_1, x_2, \ldots, x_k\}$*, and* v_{rl} *are the elements of the inverted weight matrix* W *from Theorem 1:* $V = (v_{rl}) = W^{-1}$, $det(W) \neq 0$.

Let us refer to the above estimators as the modified MM-estimator (MMM-estimator) and the modified ML-estimator (MML-estimator) respectively. It should be noted that the filtration approach (Theorem 5) is not limited to the maximum likelihood technique, it can also be used together with other known estimation methods developed for the classical (non-distorted) beta-binomial distribution. A good review of these methods can be found in (Tripathi *et al.*, 1994).

Robust estimation in the case of unknown distortion levels. Let us now consider a general case when both the BBM parameters α, β and the distortion levels ε_0, ε_1 are unknown. For simultaneous consistent estimation of α, β and $\varepsilon_0, \varepsilon_1$, two numerical algorithms are proposed; the first employs the method of moments and the second utilizes the maximum likelihood approach.

For the method of moments, the simultaneous estimation problem can be reduced to the solution of the following system of two nonlinear equations for the third and fourth order moments

$$m_3^* = m_3(\alpha(\varepsilon_0,\varepsilon_1),\beta(\varepsilon_0,\varepsilon_1),\varepsilon_0,\varepsilon_1), \quad m_4^* = m_4(\alpha(\varepsilon_0,\varepsilon_1),\beta(\varepsilon_0,\varepsilon_1),\varepsilon_0,\varepsilon_1), \tag{11}$$

where the functions $\alpha(\varepsilon_0,\varepsilon_1)$, $\beta(\varepsilon_0,\varepsilon_1)$ are expressed explicitly (see Theorem 4) from the equations for the first and second order moments

$$m_1^* = m_1(\alpha,\beta,\varepsilon_0,\varepsilon_1), \quad m_2^* = m_2(\alpha,\beta,\varepsilon_0,\varepsilon_1). \tag{12}$$

Here $m_r^* = k^{-1}\sum_{i=0}^{k} x_i^r$, $r = 1,2,3,4$; while $m_r(\alpha,\beta,\varepsilon_0,\varepsilon_1)$ are the corresponding theoretical moments for the DBBD with the parameters $n,\alpha,\beta,\varepsilon_0,\varepsilon_1$ that can be computed using Theorem 1. To solve the equations (11), let us apply the modified Newton method. Denote by J_0^c the 2x2 Jacobi matrix of the system (11) on the condition that the equations for the first two moments (12) hold. Then the iterative procedure for the solution of (11) is expressed as

$$\begin{pmatrix} \varepsilon_0^{t+1} \\ \varepsilon_1^{t+1} \end{pmatrix} = \begin{pmatrix} \varepsilon_0^{t} \\ \varepsilon_1^{t} \end{pmatrix} + \lambda \cdot (J_0^c)^{-1} \begin{pmatrix} m_3^* - m_3(\alpha(\varepsilon_0^t,\varepsilon_1^t),\beta(\varepsilon_0^t,\varepsilon_1^t),\varepsilon_0^t,\varepsilon_1^t) \\ m_4^* - m_4(\alpha(\varepsilon_0^t,\varepsilon_1^t),\beta(\varepsilon_0^t,\varepsilon_1^t),\varepsilon_0^t,\varepsilon_1^t) \end{pmatrix}, \tag{13}$$

where $\lambda \in (0,1]$ is the algorithm parameter that ensures the convergence for large distortion levels $\varepsilon_0,\varepsilon_1$ (Demidovich and Maron 1970). All expressions required for the numerical implementation of the procedure (13) are given in the Mathematical Appendix. As follows from the numerical experiments, the usual value $\lambda = 1$ (typical for the classical Newton technique) provides poor convergence, so it is prudent to start iterations with rather low λ and gradually increase it so that it becomes close to 1 in the neighborhood of the desired solution. It can be done using the recursive sequence $\lambda_{t+1} = \lambda_t \cdot (1-\theta) + \theta$, where λ_0 and θ are the tuning parameters. During the computer simulations that will be discussed below, the authors used the following values: $\lambda_0 = 0.1$, $\theta = 0.05$. Let us refer to the estimates of the model parameters α,β and the distortion levels $\varepsilon_0,\varepsilon_1$ obtained using the procedure (13) as the MMS-estimates.

For the maximum likelihood approach, the simultaneous estimation is reduced to the following constrained maximization problem

$$l(\alpha,\beta,\varepsilon_0,\varepsilon_1) = \sum_{r=0}^{n} f_r \ln(P_r(\alpha,\beta,\varepsilon_0,\varepsilon_1)) \to \max_{\alpha,\beta,\varepsilon_0,\varepsilon_1}, \quad \alpha,\beta \in R^+, \quad \varepsilon_0,\varepsilon_1 \in [0,1], \tag{14}$$

where $\{f_0, f_1, \ldots, f_n\}$ are the frequencies for the distorted sample $\{x_1, x_2, \ldots, x_k\}$, and the explicit expressions for the distorted beta-binomial probabilities $P_r(.)$ are given in

Theorem 1. This maximization problem is solved using the modification of the steepest descent method. All the expressions required for the numerical implementation are given in Mathematical Appendix. Let us refer to the estimates of α, β and $\varepsilon_0, \varepsilon_1$ obtained from (14) as the MLS-estimates.

Computer simulations. To demonstrate the robustness of the proposed estimators of the BBM parameters, a series of four computer simulations was done. It was assumed that the true values of the model parameters were $\alpha^0 = 0.5$, $\beta^0 = 9.5$, $n = 10$. These values are typical for the application area that the authors deal with (see Section 6).

Experiment 1. This experiment was dedicated to assessing the sensitivity of the beta-binomial distribution to the distortions (Theorem 1). There were generated $k = 1000$ realizations of the random variable from the DBBD with the parameters n, α^0, β^0 and the distortion levels $\varepsilon_0 = 0.01$, $\varepsilon_1 = 0.02$. For the generated sample, there were computed the empirical probabilities P_r^*, $r = 0, 1, \ldots, n$, as well as the sample mean and variance. Also, there were calculated the weight matrix W, the theoretical probabilities P_r and P_r^0, the approximate values P_r^a for P_r (the asymptotic expansion (6)), and the theoretical mean and variance for the BBD and DBBD.

As follows from the experiment results (Tables 1-3), the original beta-binomial distribution is quite sensitive to the distortions. For example, the relative difference between the non-distorted P_r^0 and distorted P_r probabilities can go up to 24.9%, and the mathematical expectation and variance can differ by 17.0% and 3% respectively. The corresponding weight matrix W (see Table 3) has the dominated leading diagonal and the adjacent elements, that explains why the linearized expressions (6) provide an accurate enough approximation of the probabilities P_r. This result validates using of stochastic expansions for assessing the sensitivity of the classical estimation and prediction techniques with respect to the distortion levels.

***Table 1**: Comparison of the original, distorted and empirical mean and variance.*

Distribution type	Mean	Variance
Classical beta-binomial distribution	0.500	0.929
Distorted beta-binomial distribution	0.585	0.957
Empirical distribution	0.577	0.943

***Table 2**: Comparison of the original, distorted and empirical probabilities for the BBM.*

r	0	1	2	3	4	5	6	7	8	9	10
	$\times10^{-1}$	$\times10^{-1}$	$\times10^{-2}$	$\times10^{-2}$	$\times10^{-2}$	$\times10^{-3}$	$\times10^{-3}$	$\times10^{-4}$	$\times10^{-4}$	$\times10^{-5}$	$\times10^{-6}$
P_r^0	6.93	1.87	7.23	2.92	1.15	4.30	1.46	4.34	1.06	1.91	1.91
P_r	6.30	2.34	8.40	3.24	1.25	4.54	1.50	4.36	1.04	1.80	1.73
P_r^a	6.28	2.39	8.22	3.21	1.24	4.52	1.50	4.35	1.03	1.80	1.72
P_r^*	6.32	2.34	8.33	3.19	1.17	5.00	0.95	2.50	1.50	0.00	1.91

Table 3: *Elements of the weight matrix W for the distortion levels* ε_0 = 0.01, ε_1 = 0.02.

	0	1	2	3	4	5	6	7	8	9	10
0	0.9044	0.0183	0.0004	$\sim10^{-5}$	$\sim10^{-7}$	$\sim10^{-9}$	$\sim10^{-10}$	$\sim10^{-12}$	$\sim10^{-14}$	$\sim10^{-15}$	$\sim10^{-17}$
1	0.0914	0.8969	0.0362	0.0011	$\sim10^{-5}$	$\sim10^{-6}$	$\sim10^{-8}$	$\sim10^{-10}$	$\sim10^{-11}$	$\sim10^{-13}$	$\sim10^{-14}$
2	0.0042	0.0815	0.8891	0.0538	0.0022	0.0001	$\sim10^{-6}$	$\sim10^{-7}$	$\sim10^{-9}$	$\sim10^{-11}$	$\sim10^{-12}$
3	0.0001	0.0033	0.0717	0.8811	0.0710	0.0036	0.0001	$\sim10^{-5}$	$\sim10^{-7}$	$\sim10^{-8}$	$\sim10^{-10}$
4	$\sim10^{-6}$	0.0001	0.0025	0.0621	0.8727	0.0879	0.0053	0.0003	$\sim10^{-5}$	$\sim10^{-7}$	$\sim10^{-8}$
5	$\sim10^{-8}$	$\sim10^{-6}$	0.0001	0.0019	0.0527	0.8641	0.1044	0.0074	0.0004	$\sim10^{-5}$	$\sim10^{-6}$
6	$\sim10^{-10}$	$\sim10^{-8}$	$\sim10^{-6}$	$\sim10^{-5}$	0.0013	0.0435	0.8551	0.1206	0.0097	0.0006	$\sim10^{-5}$
7	$\sim10^{-12}$	$\sim10^{-10}$	$\sim10^{-8}$	$\sim10^{-7}$	$\sim10^{-5}$	0.0009	0.0344	0.8460	0.1363	0.0124	0.0008
8	$\sim10^{-15}$	$\sim10^{-13}$	$\sim10^{-11}$	$\sim10^{-9}$	$\sim10^{-7}$	$\sim10^{-5}$	0.0005	0.0256	0.8366	0.1517	0.0153
9	$\sim10^{-17}$	$\sim10^{-15}$	$\sim10^{-13}$	$\sim10^{-11}$	$\sim10^{-9}$	$\sim10^{-8}$	$\sim10^{-6}$	0.0003	0.0169	0.8269	0.1667
10	$\sim10^{-20}$	$\sim10^{-18}$	$\sim10^{-16}$	$\sim10^{-14}$	$\sim10^{-12}$	$\sim10^{-10}$	$\sim10^{-8}$	$\sim10^{-6}$	0.0001	0.0083	0.8171

Experiment 2. This experiment was devoted to assessing the bias of the classical BBM parameter estimators that ignore the misclassifications (Theorems 2, 3). There were generated 100 independent random samples of size k = 1000 from the BBM with the parameters n, α^0, β^0. It was assumed that $\varepsilon_0 = \varepsilon_1 \in [\,0;\ 0.02\,]$ and they varied with the step 0.002, and each sample was contaminated according to the distortion model (2), (3). For each distorted sample and for each value of the distortion level, the classical MM and ML methods were applied. Then, for all values of $\varepsilon_0, \varepsilon_1$, the 95%-confidence intervals of the α, β estimates were computed (using the common technique, which assumes that the estimates follow the normal distribution). Finally, for the same distortion levels, the theoretical biases were obtained using the stochastic expansions (7), (8).

The results of the experiment are presented in Figure 1, where $\rho(.)$ is the relative bias (i.e. $\Delta\alpha/\alpha^0$ or $\Delta\beta/\beta^0$). As follows from the figure, the stochastic expansions (7), (8) provide good approximation of the parameters biases caused by the distortions with the levels $\varepsilon_0, \varepsilon_1 \le 0.01$. Besides, the classical estimators are quite sensitive to the distortions. For example, for the distortion levels $\varepsilon_0 = \varepsilon_1 = 0.01$, the relative errors for the parameters α, β are respectively 50%, 22.7% for the the MM-estimator and 52.7%, 28.0% for the ML-estimator.

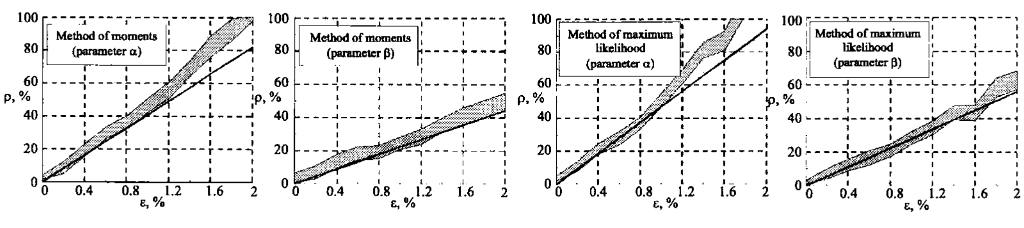


Figure 1: *The biases of the classical MM- and ML-estimators of the BBM parameters, which ignore the misclassifications: gray tubes –experimental 95% confidence intervals; solid lines– approximation via the asymptotic expansions (7), (8);* ρ *–the relative bias,* ε*– the distortion level* (ε_0=ε_1).

However, for practical applications, it is also important to analyze the sensitivity of another BBM parametrization (Prentice 1986): $\pi = \alpha/(\alpha+\beta)$, $\gamma = 1/(\alpha+\beta)$, where π is the average response probability, and γ is a measure of the inter-group correlation. For this parametrization, the relative errors of the MM and ML-estimators for the parameters π, γ are 20.9%, 19.4% and 17.7%, 22.6% respectively. It means that ignoring response misclassifications leads to quite large errors when assessing both the average response probability for the clusters and the inter-group correlation between units. This numerical result emphasizes the importance of the research topic and motivates development of new robust estimators, which take the distortion model into account.

It should be noted that Neuhaus (1999, 2002) performed similar computer simulations for the binary regression, as well as for the population-averaged and the mixed-effects logistic models. In his simulation, Neuhaus was interested in the bias of the regression coefficients and made a conclusion that the biases due to response misclassifications were negligible for small values of the distortion levels and were substantial only when $\varepsilon_0, \varepsilon_1 \geq 0.10$. Since our experiments yielded qualitatively different results (see Figure 1), this fact should be explained in details.

For the comparison purposes, the beta-mixed hierarchical model considered in this paper (both BBM and BLM) can be reformulated as a special case of the generalized linear mixed model (GLMM), which is an extension of the generalized linear model (GLM) to the longitudinal or clustered data case. The reformulation can be done by introducing dummy constant covariates for each cluster/unit, and choosing an appropriate link function and a random effects distribution. Then the regression coefficients can be considered as the beta-mixed hierarchical model parameters, and their sensitivity to the distortions can be investigated using technique employed in this paper. Hence, the above model conversion can be treated as a specific nonlinear re-parametrization of the beta-mixed hierarchical model, which leads to completely different meaning of the model parameters.

For this re-parametrization, the parameter estimator sensitivity w.r.t. the misclassifications may increase, depending on the true values of the parameter. For instance, for small values of π (which are typical for our application area), the misclassifications essentially influence the estimate $\hat{\pi}$, since $E\{x_i/n\} = \varepsilon_0(1-\pi)+(1-\varepsilon_1)\pi$. Thus, when $\pi = 0.05$ and $\varepsilon_0 = \varepsilon_1 = 0.01$ the expectation of x_i/n is equal to 0.059, i.e. misclassifications cause 18% increase of the corresponding parameter value. This justifies the qualitative difference of the Neuhaus' and ours simulation results.

Therefore, the obtained results show that the beta-binomial model parameter estimators are less robust to the response misclassifications compared to the estimators for the models investigated by Neuhaus. This emphasizes the research topic importance and motivates development of robust estimators for the BBM. It should be also noted that the robust estimation approach for the logistic-normal model that was employed by Neuhaus (2002) can not be applied to the BBM since he used specific properties of the logistic link function that the beta-binomial distribution does not possess.

Experiment 3. This experiment was aimed at the performance evaluation of the proposed robust estimators in the case of known distortion levels (Theorems 4, 5). It was assumed that $\varepsilon_0 = \varepsilon_1 = 0.01$, and the developed MMM and MML-estimators were compared to the classical MM and ML-estimators by assessing the biases, standard deviations, and histograms. As follows from the experiment results (Figure 2), the proposed estimation methods allow essentially decreasing the bias of the α, β estimates and lead to the smaller standard deviation while compared to the classical estimators. In particular, the MMM-estimator yields the relative biases 2.0%, 2.1% for the parameters α, β respectively against 47.7%, 25.2% obtained by applying the classical MM technique. The MML-estimator ensures the relative biases 0.9%, 1.1% in contrast to 54.2%, 30.3% for the classical ML method. These results confirm the robust performance of the proposed estimators.

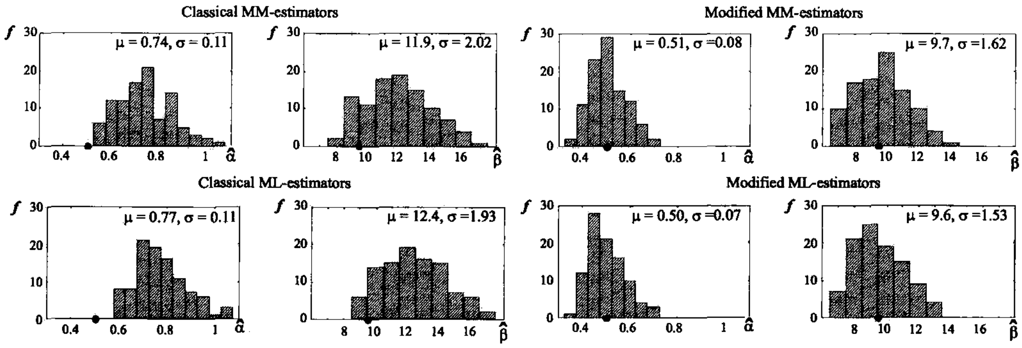


***Figure 2**: Histograms of the classical and proposed estimators of the BBM parameters for known distortion levels: f –empirical frequency, μ– sample mean, σ –sample standard deviation; the circles denote the true parameter values.*

Experiment 4. This experiment focused on the performance evaluation of the proposed robust estimators in the case of unknown distortion levels. It was assumed that $\varepsilon_0 = \varepsilon_1 = 0.01$, and the developed MMS and MLS-estimators were compared to the classical MM and ML-estimators by assessing the biases and standard deviations. As follows from the experiment results (Table), the proposed estimation techniques allow essentially decreasing the bias of the α, β estimates, while the standard deviation increases compared to the classical estimators. In particular, the MMS-estimator yields the relative biases 2.0%, 3.2% for the parameters α, β respectively against 46.0%, 24.3% obtained by applying the classical MM technique. The MLS-estimator ensures the relative biases 6.0%, 1.2% in contrast to 52.0%, 29.3% for the classical ML method. On the other hand, the standard deviation increases up to twice compared to the classical methods that ignore the misclassifications. This effect is caused by the identification of two extra parameters $\varepsilon_0, \varepsilon_1$ in addition to α, β that normally leads to extra variation.

Advantages of the developed methods were also confirmed by additional numerical research aimed at the identifiability analysis, which was based on computing of the determinant and condition number for the relevant Jacobi matrices. For the MMS-estimator, there were examined both the full 4×4 Jacobian of the system (11), (12) and the reduced 2×2 Jacobian, which is used in the numerical procedure (13). During the simulation, the determinant of the full Jacobian was far from zero and varied from 0.01 to 0.10 that confirms the identifiability. However, the corresponding condition number was rather high (from $5.8 \cdot 10^5$ to $1.4 \cdot 10^6$), that validates using of the proposed iterative procedure (13), which employs inversion of the 2×2 matrix with much better condition number (from 55.4 to 73.8). For the MLS-estimator, there was examined the 4x4 matrix of the second derivatives for the log-likelihood function (14). Its determinant was greater then 10^5 that indicates the identifiability of all model parameters. But the corresponding condition number varied from $5.2 \cdot 10^5$ to $7.8 \cdot 10^8$ that explains slow convergence of the optimization routine (approximately 85 times slower then for the MMS-estimator) due to the ravine structure of the objective function. Nevertheless, the MLS technique gives better estimation results in comparison with the MMS (in 48% of simulation runs, the MLS biases were smaller then the MMS biases for all four parameters, in 27% of runs –for three parameters, in 19% of runs– for two parameters, in 5% of runs –for one parameter, and only in 1% of runs the MMS dominated over the MLS for all the parameters). These results confirm both the identifiability and the robust performance of the developed estimators.

***Table 4**: Comparison of the classical and proposed estimators of the BBM for unknown distortion levels.*

Parameter	α (true value 0.5)				β (true value 9.5)			
Method	MM	ML	MMS	MLS	MM	ML	MMS	MLS
Mean	0.73	0.76	0.49	0.53	11.81	12.28	9.20	9.39
Standard deviation	0.12	0.11	0.21	0.22	2.01	1.97	2.63	0.71

4 Robust estimation of the beta-logistic model

Robustness of the classical ML-estimator. Let a^0, b^0 be the true unknown values of the BLM parameters, and let $\Delta\tilde{a}(\varepsilon_0, \varepsilon_1)$, $\Delta\tilde{b}(\varepsilon_0, \varepsilon_1)$ be the biases of the parameter estimators that ignore the misclassifications with the levels $\varepsilon_0, \varepsilon_1$. The following theorems evaluate the robustness of the classical ML-estimator via its bias w.r.t. the distortion levels.

Theorem 6 *The bias of the classical ML-estimator of the BLM parameters, which ignores the misclassifications, satisfies the following asymptotic expansion*

$$\begin{pmatrix} \Delta\tilde{a} \\ \Delta\tilde{b} \end{pmatrix} = -H^{-1}G \cdot \begin{pmatrix} \varepsilon_0 \\ \varepsilon_1 \end{pmatrix} + \mathbf{1}_{2m}\left(o(\varepsilon_0, \varepsilon_1) + O_P\left(\frac{1}{\sqrt{k}}\right)\right), \tag{15}$$

under the assumption that the covariates Z_i belong to the countable set $\{\vartheta_1, \vartheta_2, \ldots, \vartheta_d\} \subset R^m, i = 1, 2, \ldots, k$, all vectors $\{\vartheta_q\}$ are equiprobable, and the clusters *with the same covariates ϑ_q have equal group sizes $\check{n}_q$, where $\mathbf{1}_{2m}$ is a vector of ones of size 2m, and H, G are $(2m \times 2m)$, $(2m \times 2)$ matrices given in Mathematical Appendix.*

As follows from the theorem, the classical ML-estimator of the BLM parameters becomes biased and inconsistent under the distortions. Expression (15) allows assessing the sensitivity of this estimator to the misclassifications (3). Let us now propose consistent and unbiased estimators that take into account the distortion model (2).

Robust estimation in the case of known distortion levels. Consider the case when the distortion levels $\varepsilon_0, \varepsilon_1$ are known a priori. First, let us obtain a stochastic expansion for the biases that differs from (15) by taking into account an observed sample $X = \{x_1, x_2, \ldots, x_k\}$.

Theorem 7 *For the observed sample X, the bias of the classical ML-estimator of the BLM parameters, which ignores the misclassifications, satisfies the following asymptotic expansion*

$$\begin{pmatrix} \Delta\tilde{a} \\ \Delta\tilde{b} \end{pmatrix} = J^{-1}(a^0, b^0, X) \cdot g_e(a^0, b^0, X, \varepsilon_0, \varepsilon_1) + \mathbf{1}_{2m}\left(o(\varepsilon_0, \varepsilon_1) + O_P\left(\frac{1}{\sqrt{k}}\right)\right), \qquad (16)$$

where the $(m \times m)$-matrix $J(.)$ and the m-vector $g_e(.)$ are defined in Mathematical Appendix.

Then, the expansion (16) allows constructing a bias compensating procedure for the classical ML-estimator

$$(\tilde{a}^{t+1}, \tilde{b}^{t+1})^T = (\tilde{a}^t, \tilde{b}^t)^T - \lambda \cdot J^{-1}(\tilde{a}^t, \tilde{b}^t, X) \cdot g_e(\tilde{a}^t, \tilde{b}^t, X, \varepsilon_0, \varepsilon_1), \qquad (17)$$

where t is the iteration number and λ is the algorithm parameter that ensures convergence. The parameter λ is selected in the similar way as for the numerical procedure (13) from Section 3: $\lambda_{t+1} = \lambda_t \cdot (1-\theta) + \theta$. In the given below computer simulations, the authors used values $\lambda_0 = 0.1$, $\theta = 0.05$. Let us refer to the bias-corrected ML-estimator (17) as the modified ML-estimators (MML).

Robust estimation in the case of unknown distortion levels. Let us now consider a general case when both the BLM parameters a, b and the distortions levels ε_0, ε_1 are unknown. For simultaneous consistent estimation of a, b and $\varepsilon_0, \varepsilon_1$, a maximum likelihood based numerical algorithm is proposed.

Using results from Section 3, the log-likelihood function for the BLM that accommodates the distortion model (2), (3) may be expressed as

$$l_\varepsilon(a, b, X, \varepsilon_0, \varepsilon_1) = \sum_{i=1}^{k} \log\left(\sum_{j=0}^{n_i} w^i_{x_i j}(\varepsilon_0, \varepsilon_1) \cdot P^i_j(a, b)\right), \qquad (18)$$

where

$$P^i_j(a,b)=\binom{n_i}{j}\frac{B(\alpha_i(a)+j,\beta_i(b)+n_i-j)}{B(\alpha_i(a),\beta_i(b))},\quad \alpha_i(a)=\exp(Z_i^T a),\quad \beta_i(b)=\exp(Z_i^T b),$$

$B(.)$ is the complete beta function, and $w^i_{x_i j}$ are the weights of the distorted beta-binomial distribution with the parameters $n_i, \alpha_i(a), \beta_i(b), \varepsilon_0, \varepsilon_1$ (see Theorem 1). Then, the simultaneous estimation of the BLM parameters and the distortion levels is reduced to the following constrained maximization problem

$$l_\varepsilon(a,b,X,\varepsilon_0,\varepsilon_1)\to \max_{a,b,\varepsilon_0,\varepsilon_1},\quad a,b\in R^m,\quad \varepsilon_0,\varepsilon_1\in[0,1]. \qquad (19)$$

The problem (19) is solved using the gradient descent method; all the required expressions are given in the Mathematical Appendix. Let us refer to the estimates of a, b and $\varepsilon_0, \varepsilon_1$ obtained from (19) as the MLS-estimates.

Computer simulations. To demonstrate the robust performance of the developed methods for the estimation of the BLM, a number of computer simulations was done. It was assumed that the true values of the parameters were $a^0=1$, $b^0=2$, $\forall i, n_i=10$, $k=1000$, and the covariates $Z_i \in R$ were uniformly distributed on the segment [1.0; 1.1]. This range of the covariates corresponds to the intervals $\alpha \in [2.7; 3.0]$, $\beta \in [7.4; 9.0]$ that is typical for the application area the authors deal with (see Application Example). The simulations included three experiments.

Experiment 1. This experiment was devoted to assessing the bias of the classical ML-estimator of the BLM parameters that ignores the misclassifications (Theorem 6). There were generated 100 independent random samples of size $k = 1000$ from the BLM with the parameters a^0, b^0. It was assumed that $\varepsilon_0=\varepsilon_1\in[0;\ 0.05]$ and they varied with the step 0.01, and each sample was contaminated according to the distortion model (2), (3). For each distorted sample and for each value of the distortion level, the classical ML method was applied. Then, for all values of $\varepsilon_0, \varepsilon_1$, the 95%-confidence intervals of the a, b estimates were computed (assuming that the estimates follow the normal distribution). Finally, for the same distortion levels, the theoretical biases were obtained using the stochastic expansion (15).

As follows from the experiment results (Figure 3), the stochastic expansion (15) provides good approximation of the parameters biases caused by the distortions with the levels $\varepsilon_0, \varepsilon_1 \le 0.05$. Besides, the classical ML-estimator is quite sensitive to the distortions. For example, for the distortion levels $\varepsilon_0 = \varepsilon_1 = 0.05$, the relative errors for the parameters a, b are 39.2%, 12.1% respectively. It should be noted that the higher parameter biases in comparison with the results of Neuhaus (1999) for the binary regression are due to the specific nonlinear parametrization of the beta-mixed hierarchical models (for details, see the above discussion in the Computer Simulation subsection for the BBM).

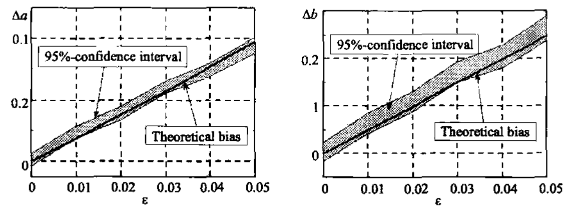


Figure 3: *The biases of the classical ML-estimator of the BLM parameters, which ignores the misclassifications: gray tubes –experimental 95% confidence intervals; solid lines– approximation via the expansion (15); $\Delta a, \Delta b$ –the relative bias, ε– the distortion level (ε_0=ε_1).*

Experiment 2. This experiment was aimed at the performance evaluation of the proposed robust estimator in the case of known distortion levels (Theorem 7). It was assumed that $\varepsilon_0 = \varepsilon_1 = 0.03$, and the developed bias-corrected estimator was compared to the classical ML-estimator by assessing the biases, standard deviations, and histograms. As follows from the experiment results (Figure 4), the proposed estimation method allows essentially decreasing the bias of the a, b estimates and leads to the similar standard deviations. In particular, the bias-corrected estimator yields the relative biases 2.3%, 1.1% for the parameters a, b respectively against 23.1%, 8.2% obtained by applying the classical ML technique.

The identifiability of the model parameters a, b and convergence of the numerical procedure (17) are determined by the properties of the 2×2 matrix of the second derivatives J for the BLM log-likelihood function (which does not take into account the distortion model). Additional numerical research indicated that the determinant of this matrix was greater then 10^5, while the corresponding condition number varied from 38.4 to 51.7. These results confirm both the identifiability and the robust performance of the bias-corrected estimator.

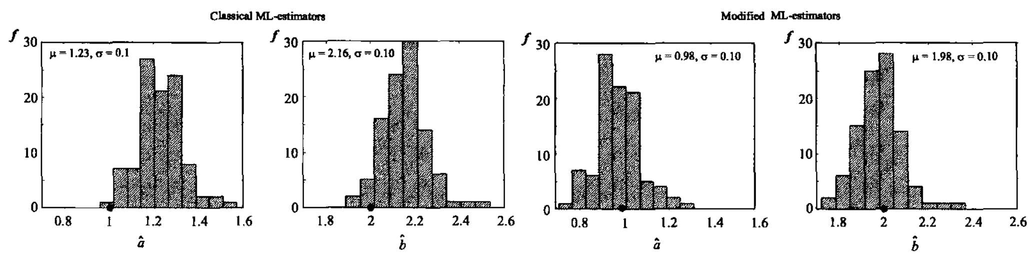


Figure 4: *Histograms of the classical and proposed estimators of the BLM parameters for known distortion levels: f –empirical frequency, μ– sample mean, σ –sample standard deviation; the circles denote the true parameter values.*

Experiment 3. This experiment focused on the performance evaluation of the proposed robust estimator in the case of unknown distortion levels. It was assumed that $\varepsilon_0 = \varepsilon_1 = 0.03$, and the developed MLS-estimator was compared to the classical ML-estimator by assessing the biases and standard deviations. As follows from the experiment results (Table 5), the proposed estimation technique allows essentially decreasing the bias of the a, b estimates, while the standard deviation is approximately the same in all cases. In particular, the MLS-estimator yields the relative biases 4.1%, 1.6% for the parameters a, b respectively against 23.0%, 7.5% obtained by applying the classical ML technique.

To analyze the identifiability of the parameters $a, b, \varepsilon_0, \varepsilon_1$ and the convergence of the developed MLS estimation algorithm, there was examined the 4×4 matrix of the second derivatives for the log-likelihood function (18), which takes into account the distortion model. Its determinant was greater then 10^7 that indicates the identifiability of all model parameters. But the corresponding condition number varied from $1.3\cdot 10^3$ to $5.1\cdot 10^4$ that explains relatively slow convergence of the optimization routine due to the ravine structure of the objective function. However, the computing time is acceptable for practical applications. These results confirm both the identifiability and the robust performance of the developed MLS-estimator.

Table 5: *Comparison of the classical and proposed estimators of the BLM for unknown distortion levels.*

Parameter	a (true value 1.0)		b (true value 2.0)	
Method	ML	MLS	ML	MLS
Mean	1.23	1.04	2.15	2.03
Standard deviation	0.11	0.12	0.11	0.11

5 Robust forecasting for beta-mixed hierarchical models

Robustness of the classical Bayes predictor. First, let us analyze the robustness of the classical Bayes predictor (1), which incorporates the true values of the model parameters α_i^0, β_i^0, assuming that the predictor input $x_i = \sum_{j=1}^{n_i} \tilde{B}_{ij}$ is contaminated by the distortions with known levels ε_0, ε_1 (here, the subscript i denotes the index of the cluster, for which the forecast is performed). The following theorem evaluates the robustness of the classical Bayes predictor w.r.t. the distortion levels by assessing the increase of the mean square error of forecasting.

Theorem 8 *If the classical Bayes predictor (1) uses the true model parameters α_i^0, β_i^0, then the mean square error of the forecast, which is based in the misclassified responses, is expressed as*

$$\tilde{r}_i^2 = r_{0i}^2 + \frac{n_i(\beta_i^0\varepsilon_0 + \alpha_i^0\varepsilon_1) + n_i^{[2-1]}((\beta_i^0)^{[2+1]}\varepsilon_0^2 - 2\alpha_i^0\beta_i^0\varepsilon_0\varepsilon_1 + (\alpha_i^0)^{[2+1]}\varepsilon_1^2)}{(\alpha_i^0+\beta_i^0)^{[2+1]}(\alpha_i^0+\beta_i^0+n_i)^2}. \quad (20)$$

where r_{0i}^2 is the error in the non-distorted case ($\varepsilon_0 = \varepsilon_1 = 0$)

$$r_{0i}^2 = \frac{\alpha_i^0\beta_i^0}{(\alpha_i^0+\beta_i^0)^{[2+1]}(\alpha_i^0+\beta_i^0+n_i)}.$$

Then, let us consider the case when the true values of the parameters α_i^0, β_i^0 are unknown, so their estimates $\hat{\alpha}_i$, $\hat{\beta}_i$ (biased because of the distortions) are used for the prediction. The following theorem evaluates the robustness of the classical Bayes predictor in this case.

Theorem 9 *If the classical Bayes predictor (1) uses the biased estimates $\hat{\alpha}_i$, $\hat{\beta}_i$ of the model parameters, then the mean square error of the forecast, which is based in the misclassified responses, is expressed as*

$$\tilde{r}_i^2 = \tilde{r}_{0i}^2 + \frac{n_i(\beta_i^0\eta_i^\beta\cdot\varepsilon_0 + \alpha_i^0\eta_i^\alpha\cdot\varepsilon_1)}{(\alpha_i^0+\beta_i^0)(\hat{\alpha}_i+\hat{\beta}_i+n_i)^2} + \frac{n_i^{[2-]}(\beta_i^{0[2+]}\cdot\varepsilon_0^2 - 2\alpha_i^0\beta_i^0\cdot\varepsilon_0\varepsilon_1 + \alpha_i^{0[2+]}\cdot\varepsilon_1^2)}{(\alpha_i^0+\beta_i^0)^{[2+]}(\hat{\alpha}_i+\hat{\beta}_i+n_i)^2}, \tag{21}$$

where $\tilde{r}_{0i}^2$ is the error in the case of the non-distorted responses but the biased parameter esimates

$$\tilde{r}_{0i}^2 = \frac{n_i\alpha_i^0\beta_i^0 + \beta_i^{0[2+]}\hat{\alpha}_i^2 - 2\alpha_i^0\beta_i^0\hat{\alpha}_i\hat{\beta}_i + \alpha_i^{0[2+]}\hat{\beta}_i^2}{(\alpha_i^0+\beta_i^0)^{[2+]}(\hat{\alpha}_i+\hat{\beta}_i+n_i)^2}, \tag{22}$$

the coefficients $\eta_i^\alpha, \eta_i^\beta$ are

$$\eta_i^\alpha = 1 + 2\frac{(\alpha_i^0+1)\hat{\beta}_i - (\hat{\alpha}_i+1)\beta_i^0}{\alpha_i^0+\beta_i^0+1}, \qquad \eta_i^\beta = 1 + 2\frac{(\beta_i^0+1)\hat{\alpha}_i - (\hat{\beta}_i+1)\alpha_i^0}{\alpha_i^0+\beta_i^0+1}, \tag{23}$$

and the ascending and descending factorials are denoted as $v^{[2+]} = v(v+1)$, $v^{[2-]} = v(v-1)$.

As follows from these theorems, the classical Bayes predictor loses its optimality under the distortions (in the sense of the mean square error of forecasting). Expressions (20), (21) allow assessing the sensitivity of the classical predictor to the misclassifications (3). Let us propose now the robust predictor that takes into account the distortion model (2).

Robust prediction under distortions. Since the results from the previous sections allow obtaining the unbiased estimates of the model parameters as well as the probability distribution of the misclassified responses, there can be derived the optimal predictor that minimizes the effect of the misclassifications in the forecast input data. This predictor is defined in the following theorem.

Theorem 10 *The optimal Bayes predictor, which takes into account the distortion model (2), (3), is expressed as the weighted sum*

$$\hat{p}_i(x) = \mathbf{E}\{p_i|x,\varepsilon_0,\varepsilon_1\} = \sum_{r=0}^{n_i}\vartheta_{xr}^i\cdot\frac{\alpha_i^0+r}{\alpha_i^0+\beta_i^0+n_i}, \tag{24}$$

where x is the sum of the distorted binary observations for the i-th cluster, and the weighting coefficients ϑ^i_{xr} *are computed from*

$$\vartheta^i_{xr} = \binom{n_i}{r} w^i_{xr} B(\alpha^0_i + r, \beta^0_i + n_i - r) \cdot \left(\sum_{l=0}^{n_i} \binom{n_i}{l} w^i_{xl} B(\alpha^0_i + l, \beta^0_i + n_i - l) \right)^{-1} \qquad (25)$$

using expressions for w^i_{sl} *given in Theorem 1.*

It can also be proved that the corresponding mean square error of forecasting is computed as

$$r^2(\hat{p}_i) = \frac{(\alpha^0_i)^{[2+]}}{(\alpha^0_i + \beta^0_i)^{[2+]}} - \sum_{x=0}^{n_i} \left(\left(\sum_{r=0}^{n_i} \vartheta^i_{xr} \frac{\alpha^0_i + r}{\alpha^0_i + \beta^0_i + n_i} \right)^2 \cdot \sum_{r=0}^{n_i} w^i_{xr} \binom{n_i}{x} \frac{(\alpha^0_i)^{[r+]} (\beta^0_i)^{[(n_i - r)+]}}{(\alpha^0_i + \beta^0_i)^{[n_i +]}} \right), \qquad (26)$$

and the p.d.f. of this forecast is

$$f_{p_i}(p|x, \varepsilon_0, \varepsilon_1) = \sum_{r=0}^{n_i} \vartheta^i_{xr} \cdot B(\alpha^0_i + r, \beta^0_i + n_i - r)^{-1} p^{\alpha^0_i + r - 1} (1 - p)^{\beta^0_i + n_i - r - 1}. \qquad (27)$$

As follows from expressions (24), (27), the proposed predictor is a weighted sum of the classical predictors for the beta-hierarchical model with shifted parameters. Also, the expression for the weights ϑ^i_{xr} are based on the Bayes formula, and ϑ^i_{xr} can be treated as the posteriori probability that the distorted value x was originated from the sum of the non-distorted binary observations r (in contrast, w^i_{xr} are the corresponding a priori probabilities). Since for the weight matrix (ϑ^i_{xr}), there can be obtained the asymptotic expansion similar to (5), it is prudent to derive an approximate expression for the proposed predictor (24), which is valid for small values of $\varepsilon_0, \varepsilon_1$.

Robust prediction for small distortion levels. If values of $\varepsilon_0, \varepsilon_1$ are small, then the sums in expressions (24), (25) can be reduced to three terms by eliminating the weighting coefficients other then $\vartheta^i_{x,x-1}$, $\vartheta^i_{x,x}$, $\vartheta^i_{x,x+1}$ for (24) and $w^i_{x,x-1}$, $w^i_{x,x}$, $w^i_{x,x+1}$ for (25). Then the robust predictor can be expressed as the classical Bayes predictor multiplied by the correction factor

$$\hat{p}_i(x) = \frac{\alpha^0_i + x}{\alpha^0_i + \beta^0_i + n_i} \cdot \frac{1 + \gamma_0 \cdot \varepsilon_0 - \gamma_1 \cdot \varepsilon_1}{1 + \xi_0 \cdot \varepsilon_0 - \xi_1 \cdot \varepsilon_1} + o(\varepsilon_0, \varepsilon_1), \qquad (28)$$

where

$$\gamma_0 = \frac{(\alpha^0_i + \beta^0_i)x - \alpha^0_i n_i}{\alpha^0_i + x}, \qquad \gamma_1 = \frac{(\alpha^0_i + \beta^0_i)x - \alpha^0_{i,+} n_i}{\beta^0_{i,-} + n_i - x},$$

$$\xi_0 = \frac{(\alpha_{i,-}^0 + \beta_i^0)x - \alpha_{i,-}^0 n_i}{\alpha_{i,-}^0 + x}, \qquad \xi_1 = \frac{(\alpha_i^0 + \beta_{i,-}^0)x - \alpha_i^0 n_i}{\beta_{i,-}^0 + n_i - x},$$

and $\alpha_{i,-}^0 = \alpha_i^0 - 1$, $\beta_{i,-}^0 = \beta_i^0 - 1$, $\alpha_{i,+}^0 = \alpha_i^0 + 1$. Expression (28) allows essentially simplifying the complexity of the robust forecasting algorithm and can be used in practical applications, for which the computing time is crucial.

Computer simulation. To demonstrate the robust performance of the developed forecasting technique, the following computer simulation was done. There was considered the beta-binomial model, and it was assumed that the true values of the model parameters were $\alpha^0=0.5$, $\beta^0=9.5$, $n=10$. For this α^0,β^0, there were generated $k=1000$ realizations of the beta random variable $p_1, p_2, \dots, p_k$ (the corresponding mean value was $\bar{p} = 0.05$). Then, for each cluster with the success probability p_i, a random Bernoulli sample of size n was obtained. Next, every sample was distorted using the expression (2) for $\varepsilon_0 = \varepsilon_1 \in [0;\ 0.05]$ varying with the step 0.01. Using these data, the ML- and MLS-estimates of the α,β parameters were computed. For each cluster, two types of the forecast was done: (i) the classical prediction (1) based on the ML-estimates, and (ii) the proposed prediction (24) based on the MLS-estimates. Finally, for every distortion level, the 95%-confidence intervals of the mean square error of forecasting were computed for the both predictors (assuming that the errors follow the normal distribution).

As follows from the experiment results (Figure 5), the developed prediction technique based on the proposed MLS-estimation algorithm ensures essentially lower mean square error of forecasting when compared to the classical estimation and prediction methods.

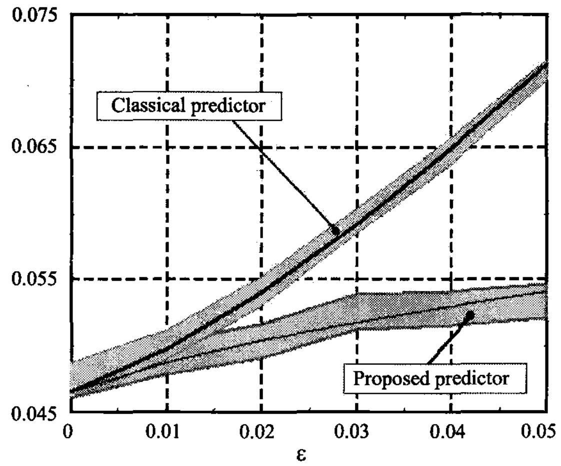


Figure 5: Comparison of the classical and proposed predictors: gray tubes – experimental 95% confidence intervals, solid lines– theoretical mean square errors of forecasting; r – mean square error, ε – distortion level (ε_0=ε_1).

For example, for the distortion levels $\varepsilon_0 = \varepsilon_1 = 0.05$, the classical procedures lead to the error $r = 0.071$ against $r_0 = 0.047$ for the non-distorted case, while the proposed robust methods ensure the error $r = 0.054$ (note that Figure 5 shows r against ε, while the above expressions are given for r^2). Hence, for the average response probability $\bar{p} = 0.05$, the increment of the forecast error $\Delta r = r - r_0$ reduces from 0.026 to 0.007. It should be stressed that the increment Δr caused the misclassifications can not be compensated completely (as follows from the Bayes forecasting theory), but the obtained value $r = 0.054$ is the lowest for these model parameters and distortion levels. These results confirm the robust performance of the developed estimation and forecasting techniques.

6 Application example

The developed methods of robust estimation and prediction were used for forecasting TV audience behaviour. This problem arises in mediaplanning (Sissors and Lincoln 1994), which focuses on optimizing of advertising schedules taking into account the target consumer groups (defined by age, sex, income, etc.) and budget constraints. For this application area, statistical forecasting of future audience behaviour using records from the past is a key issue, since it defines efficiency of the advertising spending.

Grouped binary data in mediaplanning. In TV mediaplanning, the binary responses arise as a result of exposing advertising commercials to a part of TV audience (the representative sample of the target group) during predefined TV breaks, where 1 means that a person saw the commercial and vice versa. These data are registered by special electronic devices (people-meters) and are grouped in a natural way with respect to every person and break type (defined by week day, day time, adjoining program genres, etc.).

The misclassifications that may contaminate these data are caused by improper use of the people-meters, which automatically register a TV channel being viewed, but require manual registration of household members watching the TV. It is obvious that there exists a small probability of using a wrong registration button that leads to distortions of the recorded observations. The statistical properties of the viewing data are traditionally described by the beta-binomial model (Danaher 1992), while the misclassification effect is usually ignored.

In frames of the paper notation, the TV viewing data may be interpreted as follows: $\tilde{B}_{ij}$ is the i-th person response to j-th commercial break of the certain type, k is the number of persons in a target group, and n_i is the number of the breaks that the i-th person was exposed to. It assumed that the target group and break type uniquely define the covariates Z_i, and each person's viewing behaviour for this break type is described by the success probability p_i that follows the beta distribution with the parameter α_i^0, β_i^0.

For the case studies below, there were examined two data sets for one of the German TV channels for the year 2000. The first of them focuses on improving the model adequacy, while the second one deals with increasing the forecasting accuracy.

Viewing data modelling. To demonstrate the advantages of the developed distorted beta-binomial model (DBBM), which takes into account the misclassifications and employs the proposed robust estimation techniques, there were considered the TV viewing data for eleven commercial breaks ($n = 11$) corresponding to "World News" showed on Saturday prime time. There were investigated six target groups with different sex (M,W) and age (14-29, 30-49, 50+) with size k varying from 1025 to 2488.

The results of the model adequacy analysis are presented in Table 6, which shows that the proposed DBBM and the relevant robust estimation algorithms significantly

Table 6: *Adequacy analysis of the classical (BBM) and the proposed (DBBM) models for describing the TV audience behaviour using Pearson's χ^2 goodness-of-fit statistics.*

Target group	M 14-29	M 30-49	M 50+	W 14-29	W 30-49	W 50+
			Data characteristics			
Group size, k	1137	2011	2281	1025	2084	2488
Sample mean	$1.2\cdot10^{-2}$	$4.9\cdot10^{-2}$	$3.6\cdot10^{-2}$	$2.1\cdot10^{-2}$	$3.8\cdot10^{-2}$	$4.7\cdot10^{-2}$
Overdispersion	1.66	3.33	2.75	2.05	2.54	3.27
			Classical beta-binomial model (MM-estimator)			
p-value	0.41	0.96	0.01	0.82	0.10	0.05
χ^2-statistics	9.30	3.14	21.7	5.16	14.6	17.1
Parameter α	0.17	0.16	0.17	0.18	0.21	0.16
Parameter β	13.9	3.14	4.56	8.33	5.27	3.25
			Classical beta-binomial model (ML-estimator)			
p-value	0.45	0.97	0.02	0.80	0.10	0.05
χ^2-statistics	8.88	2.87	20.0	5.43	14.6	16.9
Parameter α	0.17	0.17	0.19	0.19	0.24	0.17
Parameter β	13.5	6.55	5.08	8.81	5.78	3.47
			Distorted beta-binomial model (MMS-estimator)			
p-value	0.28	0.99	0.32	0.89	0.84	0.14
χ^2-statistics	10.9	1.00	10.4	4.35	4.99	13.5
Parameter α	0.09	0.13	0.14	0.13	0.15	0.15
Parameter β	9.57	5.58	3.88	7.07	4.00	3.14
Distortion level ε_0	0.003	0.002	0.003	0.003	0.006	0.001
Distortion level ε_1	0.060	0.000	0.042	0.000	0.060	0.004
			Distorted beta-binomial model (MLS-estimator)			
p-value	0.63	0.98	0.77	0.88	0.83	0.50
χ^2-statistics	7.11	2.46	5.67	4.40	5.04	8.34
Parameter α	0.11	0.15	0.10	0.15	0.15	0.11
Parameter β	9.12	6.39	3.14	7.51	4.31	2.36
Distortion level ε_0	0.001	0.002	0.007	0.002	0.006	0.005
Distortion level ε_1	0.048	0.015	0.068	0.008	0.020	0.111

increase the modelling accuracy. For example, for the target group M 50+ (men of age 50 and older), the classical BBM yields the p-values 0.01 for the MM-estimator and 0.02 for the ML-estimator, while the proposed DBBM ensures values 0.32 and 0.77 for the MMS and MLS estimators respectively. This confirms the applicability of the paper results to the modelling of the TV audience behaviour.

Forecasting of audience behaviour. To illustrate the accuracy of the developed forecasting technique, there were considered $N_z = 31$ commercial breaks of different types exposed in December 2000 for the target group W 50+ in the frames of a single adverting campaign. Based on the past data for the similar breaks (for three preceding months, September–November, 2000), there were obtained the viewing behaviour models based on the proposed DBBD distribution. Then, using the proposed prediction method, for all persons and all breaks, there were generated the forecasts π_{iz} (the probability that the i-th person watched the break of type z). Similar forecasts were also obtained for the classical model based on the BBD.

The accuracy for the obtained forecast was evaluated using the specific performance measures adopted in mediaplanning, the *Reach* and *GRP* (Danaher 1992). The first of them, *Reach*, describes the audience fraction (within the target group), which have seen the advertising commercial at least once during the whole advertising campaign:

$$Reach = k^{-1} \sum_{i=1}^{k} \left(1 - \prod_{z=1}^{N_z} (1 - \pi_{iz}) \right).$$

The second performance measure, *GRP* (Gross Rating Points), defines the sum of the above fractions throughout the campaign (without considering the audience duplication):

$$GRP = k^{-1} \sum_{i=1}^{k} \sum_{z=1}^{N_z} \pi_{iz} \,.$$

Using these expressions, there were obtained the *Reach–GRP* curves via considering smaller advertising campaigns composed of the considered breaks (with break number from 1 to N_z). In practice, such curves are the primary tool for media-planners who use them for assessing the economical efficiency of adding extra break to the campaign.

Figure 6 compares the *Reach–GRP* curves for the BBM and DBBM-based forecasts with the real data curve calculated using the December 2000 records. As follows from the figure, the proposed forecasting technique ensures much more accurate approximation of the *Reach–GRP* relation than the classical BBM method. In particular, the maximum relative error of the *Reach–GRP* approximation using the BBM-based forecast is about 21%, while the proposed DBBM-based technique ensures the relative error less then 4.2%. This confirms the practical value of our results.

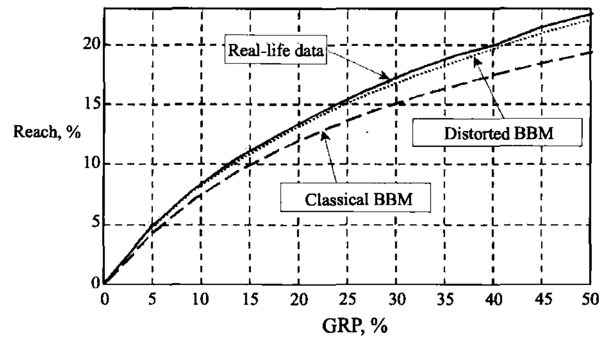


Figure 6: Comparison of the Reach–GRP curves based on the classical (BBM) and the proposed (DBBM) models against the curve obtained from the real data.

7 Conclusion

The paper proposes new robust estimation and forecasting techniques for the grouped binary data in the case of response misclassifications caused by stochastic additive distortions. It is assumed that the data are described by the beta-binomial or the beta-logistic model that belong to the class of the beta-mixed hierarchical ones. For these models, it is examined the effect of ignoring the misclassifications and there are obtained expressions for the biases of the method-of-moments and maximum likelihood estimators, as well as expressions for the increase in the mean square error for the Bayes predictor. These expressions allow assessing the sensitivity of the classical techniques w.r.t. the distortion levels and decide on their applicability in practice.

To minimize the misclassification effects, there were developed new consistent estimators and a new Bayes predictor, which take into account the distortion model. There were considered two cases (of known and unknown distortion levels), for which explicit expressions and numerical algorithms were proposed that allow constructing the small-sensitive estimators of the model parameters and the small-sensitive forecasting procedures. The robustness of the developed techniques was verified by computer simulations, and the practical value was confirmed by a real-life case study. The proposed algorithms were implemented as a MATLAB toolbox.

Future work will deal with the minimax robust estimation and forecasting for the case of known upper and lower bounds of the distortion levels, and also with the problem of small sample performance for the developed methods.

Acknowledgments This research was partly supported by the ISTC project B-705 and INTAS grant YSF 03-55-869. The authors thank Prof. Vladimir Zaiats for the financial support at the Barcelona Conference on Asymptotic Statistics, 2003. The authors are also grateful to Omega Software GmbH for the application example data and to the Reviewers whose valuable comments improved the paper.

Mathematical Appendix

Basic notation. $\mathbf{P}\{.\}$ is the probability of a random event, $\mathbf{E}\{.\}$ is the mathematical expectation of a random variable, $\mathbf{V}\{.\}$ is the variance of a random variable, $y^{[z-]} = y(y-1)\dots(y-z+1), y^{[z+]} = y(y+1)\dots(y+z-1), y \in R, z \in N$ are the incomplete factorials, $C_n^r = \binom{n}{r}$ is the binomial coefficient. Definition of O_P: for two random sequences Y_n, Z_n, $Y_n = O_P(Z_n)$ means that $\forall \epsilon > 0 \ \ \exists\, k_\epsilon, N_\epsilon$ that $0 < k_\epsilon < +\infty$, $0 < N_\epsilon < +\infty$ and for $n > N_\epsilon$, $\mathbf{P}\{|Y_n/Z_n| < k_\epsilon\} > 1 - \epsilon$.

Proof of Theorem 1. Let r be a realization of the DBBD random variable. Denote by $\{H_{rs}\}, r = 0, 1, \dots, n$, a partition complete set of disjoint events, where H_{rs} means that the distorted value r was obtained via the distortions (2) from the original positive responses count s. Then using the total probability formula

$$P_r(\varepsilon_0, \varepsilon_1) = \sum_{s=0}^{n} \mathbf{P}\{H_{rs}\} \cdot P_s^0, r = 0, 1, \dots, n.$$

To find the probability $\mathbf{P}\{H_{rs}\}$, denote by z_0, z_1 the number of the distorted zeros and ones in the original data. Then combinatorics yields to the following expression

$$w_{rs}(\varepsilon_0, \varepsilon_1) = \mathbf{P}\{H_{rs}\} = \sum_{z_0, z_1} C_n^{z_1} \varepsilon_0^{z_0} (1 - \varepsilon_0)^{n-s-z_0} \varepsilon_1^{z_1} (1 - \varepsilon_1)^{s-z_1}, \quad s - z_1 + z_0 = r.$$

Denoting $l = s + z_0 = r + z_1$ leads to $l \geq r,\ l \leq s + r,\ l \geq s,\ l \leq n$, which is equivalent to $\max(s, r) \leq l \leq \min(n, s + r)$, that proves the theorem. □

Remark. The standard approach for investigating the properties of the estimators that are fitted to the misspecified model is based on the results of White (1982) that involve Kullback-Leibler divergence. For Theorems 2, 3, 6, the authors employ a different approach that allows using the specific DBBD properties to obtain elegant proofs. However, one can check that using the Kullback-Leibler divergence leads to the exactly the same results.

Proof of Theorem 2. The classical MM-estimator of the BBM parameters is expressed as (Johnson *et al.* 1996):

$$\tilde{\alpha}_{MM} = \frac{(n - \bar{x} - s^2/\bar{x})\bar{x}}{(s^2/\bar{x} + \bar{x}/n - 1)n}, \quad \tilde{\beta}_{MM} = \frac{(n - \bar{x} - s^2/\bar{x})(n - \bar{x})}{(s^2/\bar{x} + \bar{x}/n - 1)n}, \tag{29}$$

where $\bar{x}$ is the sample average and s^2 is the sample variance. Let $m(\varepsilon_0, \varepsilon_1)$, $d(\varepsilon_0, \varepsilon_1)$ be the mean and variance of the DBBD with the parameters $n, \alpha^0, \beta^0, \varepsilon_0, \varepsilon_1$ (see Theorem 1). Since $\bar{x}$, s^2 are unbiased and consistent estimators, and $\mathbf{V}\{\bar{x}\} = d(\varepsilon_0, \varepsilon_1)/k$,

$V\{s^2\} = O(1/k)$ (Ivchenko and Medvedev 1984), then $\bar{x} = m(\varepsilon_0, \varepsilon_1) + O_P(1/\sqrt{k})$, $s^2 = d(\varepsilon_0, \varepsilon_1) + O_P(1/\sqrt{k})$. Using these expressions together with the properties of $O_P(.)$ to modify (29), we get

$$\tilde{\alpha}_{MM}(\varepsilon_0, \varepsilon_1) = \frac{(n - m(\varepsilon_0, \varepsilon_1) - d(\varepsilon_0, \varepsilon_1)/m(\varepsilon_0, \varepsilon_1))\, m(\varepsilon_0, \varepsilon_1)}{(d(\varepsilon_0, \varepsilon_1)/m(\varepsilon_0, \varepsilon_1) + m(\varepsilon_0, \varepsilon_1)/n - 1)n} + O_P(1/\sqrt{k}),$$

$$\tilde{\beta}_{MM}(\varepsilon_0, \varepsilon_1) = \frac{(n - m(\varepsilon_0, \varepsilon_1) - d(\varepsilon_0, \varepsilon_1)/m(\varepsilon_0, \varepsilon_1))\,(n - m(\varepsilon_0, \varepsilon_1))}{(d(\varepsilon_0, \varepsilon_1)/m(\varepsilon_0, \varepsilon_1) + m(\varepsilon_0, \varepsilon_1)/n - 1)\,n} + O_P(1/\sqrt{k}).$$

Employing the expressions for $m(\varepsilon_0, \varepsilon_1)$, $d(\varepsilon_0, \varepsilon_1)$ and the linear term of the Taylor expansion with the Peano remainder for the above functions of $\varepsilon_0, \varepsilon_1$ proves the theorem. □

Expressions for Theorem 3. In the theorem statement, the following notation is used:

$$P^0_{0,s} = P^0_s(\alpha^0,\beta^0), \quad P^\Sigma_s(\varepsilon_0, \varepsilon_1) = \sum_{r=0}^{s} P_r(\alpha^0,\beta^0,\varepsilon_0,\varepsilon_1), \quad s = 0, 1, \ldots, n,$$

$$S_\alpha = \sum_{s=0}^{n-1} \frac{1 - P^\Sigma_s(0,0)}{(\alpha^0 + s)^2}, \quad S_\beta = \sum_{s=0}^{n-1} \frac{P^\Sigma_s(0,0)}{(\beta^0 + n - s - 1)^2}, \quad S_{\alpha\beta} = \sum_{s=0}^{n-1} \frac{1}{(\alpha^0 + \beta^0 + s)^2},$$

$$S_{\alpha p} = -\sum_{s=0}^{n-1} \frac{(n-s)P^0_{0,s}}{\alpha^0 + s}, \quad S^+_{\alpha p} = \sum_{s=0}^{n-1} \frac{(s+1)P^0_{0,s+1}}{\alpha^0 + s}, \quad S_{\beta p} = \sum_{s=0}^{n-1} \frac{(n-s)P^0_{0,s}}{\beta^0 + n - s - 1},$$

$$S^+_{\beta p} = -\sum_{s=0}^{n-1} \frac{(s+1)P^0_{0,s+1}}{\beta^0 + n - s - 1},$$

$$H = \{H_{ij}\}_{2\times 2}, \quad G = \{G_{ij}\}_{2\times 2}, \quad H_{11} = S_{\alpha\beta} - S_\alpha,$$

$$H_{12} = H_{21} = S_{\alpha\beta}, \quad H_{22} = S_{\alpha\beta} - S_\beta, \quad G_{11} = S_{\alpha p}, \quad G_{12} = S^+_{\alpha p} = S^+_{\beta p} = G_{21}, \quad G_{22} = S_{\beta p}.$$

Proof of Theorem 3. The ML-estimator for the BBM is defined as a solution of the following system of two equations (Johnson *et al.* 1996)

$$\sum_{r=0}^{n-1} \frac{k - F_r}{\alpha + r} - \sum_{i=0}^{n-1} \frac{k}{\alpha + \beta + r} = 0, \quad \sum_{r=0}^{n-1} \frac{F_r}{\beta + n - r - 1} - \sum_{r=0}^{n-1} \frac{k}{\alpha + \beta + r} = 0, \tag{30}$$

where $F_r = f_0 + f_1 + \ldots + f_r$, and $\{f_s\}$ are the empirical frequencies. The system has a single solution that maximizes the likelihood function (Johnson *et al.* 1996). By definition, the frequencies are the binomial random variables with the parameters k, $P_s(\alpha^0,\beta^0,\varepsilon_0,\varepsilon_1)$. Since for a discrete probability distribution, the relative frequencies $\tilde{f}_s = f_s/k$ are unbiased and consistent estimators of the corresponding theoretical probabilities, and

$\mathbf{V}\{\widetilde{f_s}\} = P_s(\alpha^0,\beta^0,\varepsilon_0,\varepsilon_1)(1 - P_s(\alpha^0,\beta^0,\varepsilon_0,\varepsilon_1))/k$, then $\widetilde{f_s} = f_s/k = P_s(\alpha^0,\beta^0,\varepsilon_0,\varepsilon_1) + O_P(1/\sqrt{k})$, $s = 0,1,...,n$. As a result, the system (30) can be expressed as

$$\sum_{r=0}^{n-1}\frac{1-P_r^{\Sigma}(\varepsilon_0,\varepsilon_1)}{\alpha+r} - \sum_{r=0}^{n-1}\frac{1}{\alpha+\beta+r} + O_P\left(\frac{1}{\sqrt{k}}\right) = 0,$$

$$\sum_{r=0}^{n-1}\frac{P_r^{\Sigma}(\varepsilon_0,\varepsilon_1)}{\beta+n-r-1} - \sum_{r=0}^{n-1}\frac{1}{\alpha+\beta+r} + O_P\left(\frac{1}{\sqrt{k}}\right) = 0.$$

Let us linearize the obtained system by $\varepsilon_0,\varepsilon_1$ in the neighborhood of the point $(\alpha^0,\beta^0,0,0)$, then

$$A_\alpha^0\Delta\widetilde{\alpha}_{ML}(\varepsilon_0,\varepsilon_1) + A_\beta^0\Delta\widetilde{\beta}_{ML}(\varepsilon_0,\varepsilon_1) + A_{\varepsilon_0}^0\varepsilon_0 + A_{\varepsilon_1}^0\varepsilon_1 + o(\varepsilon_0) + o(\varepsilon_1) + O_P(1/\sqrt{k}) = 0,$$

$$B_\alpha^0\Delta\widetilde{\alpha}_{ML}(\varepsilon_0,\varepsilon_1) + B_\beta^0\Delta\widetilde{\beta}_{ML}(\varepsilon_0,\varepsilon_1) + B_{\varepsilon_0}^0\varepsilon_0 + B_{\varepsilon_1}^0\varepsilon_1 + o(\varepsilon_0) + o(\varepsilon_1) + O_P(1/\sqrt{k}) = 0,$$

where the coefficients are the corresponding derivatives. Expressing the $\Delta\widetilde{\alpha}_{ML}(\varepsilon_0,\varepsilon_1)$, $\Delta\widetilde{\beta}_{ML}(\varepsilon_0,\varepsilon_1)$ in terms of $\varepsilon_0,\varepsilon_1$ from this system proves the theorem. □

Proof of Theorem 4. Using Theorem 1, one can show that the MM-estimator of the BBM parameters α,β that takes into account the distortions model (2) is defined as a solution of the following system of two equations

$$m_1^* = n\frac{\alpha}{\alpha+\beta} + n\frac{\beta}{\alpha+\beta}\cdot\varepsilon_0 - n\frac{\alpha}{\alpha+\beta}\cdot\varepsilon_1, \tag{31}$$

$$m_2^* = m_1^* + n^{[2-]}\cdot\frac{\alpha^{[2+]} + \beta^{[2+]}\varepsilon_0^2 + \alpha^{[2+]}\varepsilon_1^2 - 2\alpha\beta\varepsilon_0 - \alpha^{[2+]}\varepsilon_1 - 2\alpha\beta\varepsilon_0\varepsilon_1}{(\alpha+\beta)^{[2+]}}. \tag{32}$$

Using the substitution

$$u = \frac{\alpha}{\alpha+\beta},\quad v = \frac{\alpha+1}{\alpha+\beta+1},\quad \alpha = \frac{u(1-v)}{v-u},\quad \beta = \frac{(1-v)(1-u)}{v-u},$$

transforms the above system into

$$m_1^* = n(u+(1-u)\varepsilon_0-u\varepsilon_1),\quad m_2^* = m_1^*+n(n-1)(vu(1-\varepsilon_0-\varepsilon_1)+\varepsilon_0^2+2u\varepsilon_0(1-\varepsilon_0-\varepsilon_1)).$$

Solving this system with respect to u, v and changing the variables back to α,β proves the theorem. □

Proof of Theorem 5. The empirical probabilities vector $\hat{P}(\varepsilon_0,\varepsilon_1)$ satisfies the following asymptotic expression (see the proof of Theorem 3): $\hat{P}_r(\varepsilon_0,\varepsilon_1) = P_r(\varepsilon_0,\varepsilon_1)+O_P(1/\sqrt{k})$,

$r = 0, 1, \ldots, n$. Using the result of Theorem 1, one gets $\hat{P}_r(\varepsilon_0, \varepsilon_1) = W(\varepsilon_0, \varepsilon_1) \cdot P^0 + O_P(1/\sqrt{k})$. Using the properties of $O_P(1/\sqrt{k})$ and the notation (10) concludes the proof. □

MMS-estimator of BBM parameters. The Jacobi matrix J_0^c for the iterative procedure (13) is calculated as $J_0^c = H \cdot G + S$, where

$$H_{11} = n^{[3-]} \frac{\alpha^{[3+]}}{(\alpha+\beta)^{[3+]}} \sum_{i=0}^{2} \left(\frac{1}{\alpha+i} - \frac{1}{\alpha+\beta+i} \right),$$

$$H_{12} = n^{[3-]} \frac{\alpha^{[3+]}}{(\alpha+\beta)^{[3+]}} \sum_{i=0}^{2} \left(\frac{-1}{\alpha+\beta+i} \right),$$

$$H_{21} = n^{[4-]} \frac{\alpha^{[4+]}}{(\alpha+\beta)^{[4+]}} \sum_{i=0}^{3} \left(\frac{1}{\alpha+i} - \frac{1}{\alpha+\beta+i} \right) + 6 \cdot H_{11},$$

$$H_{22} = n^{[4-]} \frac{\alpha^{[4+]}}{(\alpha+\beta)^{[4+]}} \sum_{i=0}^{3} \left(\frac{-1}{\alpha+\beta+i} \right) + 6 \cdot H_{12},$$

$$G_{11} = -(\alpha + 2\beta + 1), \quad G_{12} = -\alpha(\alpha+1)/\beta,$$
$$G_{21} = -\beta(\beta+1)/\alpha, \quad G_{22} = -\left(2\alpha^0 + \beta^0 + 1\right),$$

$$S_{11} = 3n^{[3-]} \frac{\alpha^{[2+]}\beta}{(\alpha+\beta)^{[3+]}}, \qquad S_{12} = 3n^{[3-]} \frac{\alpha^{[3+]}}{(\alpha+\beta)^{[3+]}},$$

$$S_{21} = 14n^{[4-]} \frac{\alpha^{[3+]}\beta}{(\alpha+\beta)^{[4+]}} + 6 \cdot S_{11}, \qquad S_{22} = 4n^{[4-]} \frac{\alpha^{[4+]}\beta}{(\alpha+\beta)^{[4+]}} + 6 \cdot S_{12}.$$

MLS-estimator of BBM parameters. The partial derivatives of the log-likelihood function $l(\alpha, \beta, \varepsilon_0, \varepsilon_1)$ are computed as

$$\frac{\partial l}{\partial \alpha} = \sum_{r=0}^{n} \left(f_r \sum_{i=0}^{n} w_{ri}(\varepsilon_0, \varepsilon_1) \cdot \frac{\partial P_i^0(\alpha, \beta)/\partial \alpha}{P_r(\alpha, \beta, \varepsilon_0, \varepsilon_1)} \right),$$

$$\frac{\partial l}{\partial \beta} = \sum_{r=0}^{n} \left(f_r \sum_{i=0}^{n} w_{ri}(\varepsilon_0, \varepsilon_1) \cdot \frac{\partial P_i^0(\alpha, \beta)/\partial \beta}{P_r(\alpha, \beta, \varepsilon_0, \varepsilon_1)} \right),$$

$$\frac{\partial l}{\partial \varepsilon_0} = \sum_{r=0}^{n} \left(f_r \sum_{i=0}^{n} \frac{\partial w_{ri}(\varepsilon_0, \varepsilon_1)}{\partial \varepsilon_0} \cdot \frac{P_i^0(\alpha, \beta)}{P_r(\alpha, \beta, \varepsilon_0, \varepsilon_1)} \right),$$

$$\frac{\partial l}{\partial \varepsilon_1} = \sum_{r=0}^{n} \left(f_r \sum_{i=0}^{n} \frac{\partial w_{ri}(\varepsilon_0, \varepsilon_1)}{\partial \varepsilon_1} \cdot \frac{P_i^0(\alpha, \beta)}{P_r(\alpha, \beta, \varepsilon_0, \varepsilon_1)} \right),$$

where

$$\frac{\partial P_i^0}{\partial \alpha} = P_i^0(\alpha,\beta) \cdot \left(\sum_{j=0}^{i-1} \frac{1}{\alpha + j} - \sum_{j=0}^{n-1} \frac{1}{\alpha + \beta + j} \right),$$

$$\frac{\partial P_i^0}{\partial \beta} = P_i^0(\alpha,\beta) \cdot \left(\sum_{j=0}^{n-i-1} \frac{1}{\beta + j} - \sum_{j=0}^{n-1} \frac{1}{\alpha + \beta + j} \right),$$

$$\frac{\partial w_{ri}}{\partial \varepsilon_0} = \sum_{l=max(i,r)}^{min(n,i+r)} C_i^{l-r} C_{n-i}^{l-i} \left((l-i)\varepsilon_0^{l-i-1}(1-\varepsilon_0)^{n-l} - (n-l)\varepsilon_0^{l-i}(1-\varepsilon_0)^{n-l-1} \right) \varepsilon_1^{l-r}(1-\varepsilon_1)^{i+r-l},$$

$$\frac{\partial w_{ri}}{\partial \varepsilon_1} = \sum_{l=max(i,r)}^{min(n,i+r)} C_i^{l-r} C_{n-i}^{l-i} \varepsilon_0^{l-i}(1-\varepsilon_0)^{n-l} \left((l-r)\varepsilon_1^{l-r-1}(1-\varepsilon_1)^{i+r-l} - (i+r-l)\varepsilon_1^{l-r}(1-\varepsilon_1)^{i+r-l-1} \right).$$

Expressions for Theorem 6. In the theorem statement, the following notation is used:

$$H_{ls} = \sum_{q=1}^{d} \vartheta_{ql}\vartheta_{qs}\check{\alpha}_q^0 \left(\sum_{j=0}^{\check{n}_q-1} \left(\frac{1-\pi_j^q}{\check{\alpha}_q^0 + j} - \frac{1}{\check{\alpha}_q^0 + \check{\beta}_q^0 + j} \right) - \sum_{j=0}^{\check{n}_q-1} \left(\frac{1-\pi_j^q}{(\check{\alpha}_q^0 + j)^2} - \frac{1}{(\check{\alpha}_q^0 + \check{\beta}_q^0 + j)^2} \right) \check{\alpha}_q^0 \right),$$

$$l, s = 1, \ldots, m,$$

$$H_{ls} = \sum_{q=1}^{d} \vartheta_{ql}\vartheta_{qs}\check{\beta}_q^0 \left(\sum_{j=0}^{\check{n}_q-1} \left(\frac{\pi_j^q}{\check{\beta}_q^0 + \check{n}_q - j - 1} - \frac{1}{\check{\alpha}_q^0 + \check{\beta}_q^0 + j} \right) - \sum_{j=0}^{\check{n}_q-1} \left(\frac{\pi_j^q}{(\check{\beta}_q^0 + \check{n}_q - j - 1)^2} - \frac{1}{(\check{\alpha}_q^0 + \check{\beta}_q^0 + j)^2} \right) \check{\beta}_q^0 \right),$$

$$l, s = m+1, \ldots, 2m,$$

$$H_{sl} = H_{ls} = \sum_{q=1}^{d} \vartheta_{ql}\vartheta_{qs}\check{\alpha}_q^0\check{\beta}_q^0 \sum_{j=0}^{\check{n}_q-1} \frac{1}{(\check{\alpha}_q^0 + \check{\beta}_q^0 + j)^2}, \quad l = 1, \ldots, m, \quad s = m+1, \ldots, 2m,$$

$$G_{l1} = \sum_{q=1}^{d} \vartheta_{ql}\check{\alpha}_q^0 \sum_{j=0}^{\check{n}_q-1} \frac{\check{n}_q - j}{\check{\alpha}_q^0 + j} \check{P}_j^q, \quad G_{l2} = -\sum_{q=1}^{d} \vartheta_{ql}\check{\alpha}_q^0 \sum_{j=0}^{\check{n}_q-1} \frac{j+1}{\check{\alpha}_q^0 + j} \check{P}_{j+1}^q, \quad l = 1, \ldots, m,$$

$$G_{l1} = -\sum_{q=1}^{d} \vartheta_{ql}\check{\beta}_q^0 \sum_{j=0}^{\check{n}_q-1} \frac{\check{n}_q - j}{\check{\beta}_q^0 + \check{n}_q - j - 1} \check{P}_j^q, \quad G_{l2} = \sum_{q=1}^{d} \vartheta_{ql}\check{\beta}_q^0 \sum_{j=0}^{\check{n}_q-1} \frac{j+1}{\check{\beta}_q^0 + \check{n}_q - j - 1} \check{P}_{j+1}^q,$$

$$l = m+1, \ldots, 2m,$$

$$\pi_j^q = \sum_{z=0}^{j} \check{P}_z^q(a^0, b^0), \quad \check{P}_j^q(a^0, b^0) = C_{\check{n}_q}^j \frac{B(\check{\alpha}_q^0 + j, \check{\beta}_q^0 + \check{n}_q - j)}{B(\check{\alpha}_q^0, \check{\beta}_q^0)}, \quad \check{\alpha}_q^0 = e^{a^{0T}\vartheta_q}, \quad \check{\beta}_q^0 = e^{b^{0T}\vartheta_q}.$$

Proof of Theorem 6. The log-likelihood function for the BLM is expressed as (Slaton et al. 2000)

$$l(a,b) = \sum_{i=1}^{k} \left(\ln(C_{n_i}^{x_i}) + \sum_{j=0}^{x_i-1} \ln(\alpha_i(a) + j) + \sum_{j=0}^{n_i-x_i-1} \ln(\beta_i(b) + j) - \sum_{j=0}^{n_i-1} \ln(\alpha_i(a) + \beta_i(b) + j) \right). \tag{33}$$

Under the theorem assumptions, the function $l(a,b)$ can be rewritten as

$$l(a,b) = \sum_{q=1}^{d}\sum_{t=1}^{k_q}\left(\ln(C_{\check{n}_q}^{y_t^q}) + \sum_{j=0}^{y_t^q-1}\ln(\check{\alpha}_q(a)+j) + \sum_{j=0}^{\check{n}_q-y_t^q-1}\ln(\check{\beta}_q(b)+j) - \sum_{j=0}^{\check{n}_q-1}\ln(\check{\alpha}_q(a)+\check{\beta}_q(b)+j)\right),$$

where k_q is a number of clusters with factors vector ϑ_q, $k = \sum_{q=1}^{d} k_q$, y_t^q is the observed number of successes for the cluster type t, $X = \bigcup_{q=1}^{d}\{y_1^q, y_2^q, \ldots, y_{k_q}^q\}$, and $\check{\alpha}_q(a) = e^{a^T\vartheta_q}$, $\check{\beta}_q(a) = e^{b^T\vartheta_q}$. Then, transforming the sum by t using approach of Johnson *et al.* (1996) for the BBM likelihood system derivation yields

$$l(a,b) = \lambda + \sum_{q=1}^{d} k_q \sum_{j=0}^{\check{n}_q-1}\Big((1-F_j^q)\cdot\ln(\check{\alpha}_q(a)+j) + F_j^q\cdot\ln(\check{\beta}(b)+\check{n}_q-j-1) - \ln(\check{\alpha}_q(a)+\check{\beta}_q(b)+j)\Big),$$

where λ is some constant, $F_j^q = \sum_{z=0}^{j} f_z^q$, and f_z^q is a relative frequency of the value z occurrence in a sample $\{y_1^q, y_2^q, \ldots, y_{k_q}^q\}$. Let us use the following asymptotic property of f_z^q (Ivchenko and Medvedev 1984): $f_z^q = \tilde{P}_z^q + O_P(1/\sqrt{k_q})$, where $\tilde{P}_z^q$ is the corresponding theoretical probability. Then, using the properties of $O_P(.)$ and the assumption that the factors $\{\vartheta_1, \vartheta_2, \ldots, \vartheta_d\}$ are equiprobable, it can be proved that for $k \to \infty$, the ML-estimator maximizes the following function

$$l_1(a,b) = \sum_{q=1}^{d}\sum_{j=0}^{\check{n}_q-1}\Big((1-\tilde{\pi}_j^q)\cdot\ln(\check{\alpha}_q(a)+j) + \tilde{\pi}_j^q\cdot\ln(\check{\beta}_q(b)+\check{n}_q-j-1) - \ln(\check{\alpha}_q(a)+\check{\beta}_q(b)+j)\Big) + O_P\left(\frac{1}{\sqrt{k}}\right),$$

where $\tilde{\pi}_j^q = \sum_{z=0}^{j}\tilde{P}_z^q(a^0, b^0, \varepsilon_0, \varepsilon_1)$, and $\tilde{P}_z^q(a^0, b^0, \varepsilon_0, \varepsilon_1)$ are the elements of the probability row for the DBBD with the parameters $\check{n}_q, \check{\alpha}_q, \check{\beta}_q, \varepsilon_0, \varepsilon_1$ (see Theorem 1):

$$\tilde{P}_z^q(a^0, b^0, \varepsilon_0, \varepsilon_1) = \sum_{l=1}^{\check{n}_q} w_{zl}^q(\varepsilon_0, \varepsilon_1)\cdot\check{P}_z^q(a^0, b^0), \quad w_{zj}^q = \sum_{l=\max(z,j)}^{\min(\check{n}_q, z+j)} C_z^{l-z} C_{\check{n}_q-j}^{l-j}\varepsilon_0^{l-j}(1-\varepsilon_0)^{\check{n}_q-l}\varepsilon_1^{l-z}(1-\varepsilon_1)^{j+z-l}.$$

Besides, it can be proved that the following asymptotic expansions for $\tilde{P}_z^q$ hold

$$\tilde{P}_z^q = \check{P}_z^q + \left((\check{n}_q-z+1)\check{P}_{z-1}^q - (\check{n}_q-z)\check{P}_z^q\right)\varepsilon_0 + \left((z+1)\check{P}_{z+1}^q - z\check{P}_z^q\right)\varepsilon_1 + o(\varepsilon_0) + o(\varepsilon_1), \quad z = 0, 1, \ldots, \check{n}_q, \quad (34)$$

where $\check{P}_{-1}^q = \check{P}_{\check{n}_q+1}^q = 0$. Since the ML-estimator is a solution of the optimization problem $l_1(a,b) \to \max$, then the corresponding partial derivable are equal to zero:

$$\sum_{q=1}^{d}\vartheta_q\check{\alpha}_q(a)\sum_{j=0}^{\check{n}_q-1}\left(\frac{1-\tilde{\pi}_j^q(a,b,\varepsilon_0,\varepsilon_1)}{\check{\alpha}_q(a)+j} - \frac{1}{\check{\alpha}_q(a)+\check{\beta}_q(b)+j}\right) + \mathbf{1}_m\cdot O_P\left(\frac{1}{\sqrt{k}}\right) = \mathbf{0}_m, \quad (35)$$

$$\sum_{q=1}^{d} \vartheta_q \breve{\beta}_q(a) \sum_{j=0}^{\breve{n}_q-1} \left(\frac{\tilde{\pi}_j^q(a,b,\varepsilon_0,\varepsilon_1)}{\breve{\beta}_q(a)+\breve{n}_q-j-1} - \frac{1}{\breve{\alpha}_q(a)+\breve{\beta}_q(b)+j} \right) + \mathbf{1}_m \cdot O_P\left(\frac{1}{\sqrt{k}}\right) = \mathbf{0}_m, \quad (36)$$

where $\mathbf{0}_m$ is a vector of zeros of size m. Linearizing this system w.r.t. $\Delta a(\varepsilon_0, \varepsilon_1)$, $\Delta b(\varepsilon_0, \varepsilon_1)$ and expressing the biases from the linearized system concludes the proof.

□

Expressions for Theorem 7. In the theorem statement, the following notation is used:

$$J = \begin{pmatrix} J^{Aa} & J^{Ab} \\ J^{Ba} & J^{Bb} \end{pmatrix}, \quad (g_\varepsilon)^T = (g^a, g^b)^T,$$

where

$$J_{ls}^{Aa} = \sum_{i=1}^{k} Z_{il} Z_{is} \alpha_i^0 \left(\sum_{j=0}^{x_i-1} \frac{1}{\alpha_i^0+j} - \sum_{j=0}^{n_i-1} \frac{1}{\alpha_i^0+\beta_i^0+j} - \alpha_i^0 \left(\sum_{j=0}^{x_i-1} \frac{1}{(\alpha_i^0+j)^2} - \sum_{j=0}^{n_i-1} \frac{1}{(\alpha_i^0+\beta_i^0+j)^2} \right) \right),$$

$$J_{ls}^{Bb} = \sum_{i=1}^{k} Z_{il} Z_{is} \beta_i^0 \left(\sum_{j=0}^{n_i-x_i-1} \frac{1}{\beta_i^0+j} - \sum_{j=0}^{n_i-1} \frac{1}{\alpha_i^0+\beta_i^0+j} - \beta_i^0 \left(\sum_{j=0}^{n_i-x_i-1} \frac{1}{(\beta_i^0+j)^2} - \sum_{j=0}^{n_i-1} \frac{1}{(\alpha_i^0+\beta_i^0+j)^2} \right) \right),$$

$$J_{ls}^{Ab} = J_{ls}^{Ba} = \sum_{i=1}^{k} Z_{il} Z_{is} \alpha_i^0 \beta_i^0 \sum_{j=0}^{n_i-1} \frac{1}{(\alpha_i^0+\beta_i^0+j)^2}, \qquad l,s = 1,2,\ldots,m.$$

$$g_l^a = \sum_{i=1}^{k} Z_{il} \alpha_i^0 \left(\frac{-x_i(\beta_i^0+n_i-x_i)}{(\alpha_i^0+x_i-1)^2} \varepsilon_0 + \frac{n_i-x_i}{\beta_i^0+n_i-x_i-1} \varepsilon_1 \right),$$

$$g_l^b = \sum_{i=1}^{k} Z_{il} \beta_i^0 \left(\frac{x_i}{\alpha_i^0+x_i-1} \varepsilon_0 + \frac{(n_i-x_i)(\alpha_i^0+x_i)}{(\beta_i^0+n_i-x_i-1)^2} \varepsilon_1 \right).$$

Proof of Theorem 7. Using the asymptotic expansion (34) and the properties of the BBD (Johnson *et al.* 1996), the log-likelihood function $l_\varepsilon(a,b,X,\varepsilon_0,\varepsilon_1)$ can be expressed in the following asymptotic form

$$l_\varepsilon(a,b,X,\varepsilon_0,\varepsilon_1) = l(a,b,X) + e(a,b,X,\varepsilon_0,\varepsilon_1) + o(\varepsilon_0) + o(\varepsilon_1),$$

where

$$e(a,b,X,\varepsilon_0,\varepsilon_1) = \sum_{i=1}^{k} \left(\left(x_i \frac{\beta_i(b)+n_i-x_i}{\alpha_i(a)+x_i-1} - (n_i-x_i) \right) \cdot \varepsilon_0 + \left((n_i-x_i) \frac{\alpha_i(a)+x_i}{\beta_i(b)+n_i-x_i-1} - x_i \right) \cdot \varepsilon_1 \right),$$

and $l(.)$ is defined by (33). Let us note that, when ignoring the distortions, the ML-estimator $\tilde{a}(X,\varepsilon_0,\varepsilon_1)$, $\tilde{b}(X,\varepsilon_0,\varepsilon_1)$ is the solution of the optimization problem $l(a,b) \to$ max. However, when taking the distortions into account, the ML-estimator $\tilde{a}^0(X,\varepsilon_0,\varepsilon_1)$, $\tilde{b}^0(X,\varepsilon_0,\varepsilon_1)$ is the solution of another problem: $l_\varepsilon(a,b) \to$ max. Let us denote

$$y=\begin{pmatrix} a \\ b \end{pmatrix},\quad \tilde{y}(X,\varepsilon_0,\varepsilon_1)=\begin{pmatrix} \tilde{a}(X,\varepsilon_0,\varepsilon_1) \\ \tilde{b}(X,\varepsilon_0,\varepsilon_1) \end{pmatrix},\quad \tilde{y}^0(X,\varepsilon_0,\varepsilon_1)=\begin{pmatrix} \tilde{a}^0(X,\varepsilon_0,\varepsilon_1) \\ \tilde{b}^0(X,\varepsilon_0,\varepsilon_1) \end{pmatrix},\quad y^0=\begin{pmatrix} a^0 \\ b^0 \end{pmatrix}.$$

It can be proved that, in the neighborhood of $\tilde{y}^0$, $\frac{\partial l}{\partial y}(\tilde{y}^0)+J(\tilde{y}^0)\cdot(\tilde{y}-\tilde{y}^0)+o(\tilde{y}-\tilde{y}^0)=\mathbf{0}_{2m}$. On the other hand, $\frac{\partial l}{\partial y}(\tilde{y}^0)=-g_e(\tilde{y}^0)+\mathbf{1}_{2m}(o(\varepsilon_0)+o(\varepsilon_1))$, where

$$\frac{\partial l(a,b,X)}{\partial a}=\sum_{i=1}^{k} Z_i\alpha_i(a)\left(\sum_{j=0}^{x_i-1}\frac{1}{\alpha_i(a)+j}-\sum_{j=0}^{n_i-1}\frac{1}{\alpha_i(a)+\beta_i(a)+j}\right), \tag{37}$$

$$\frac{\partial l(a,b,X)}{\partial b}=\sum_{i=1}^{k} Z_i\beta_i(a)\left(\sum_{j=0}^{n_i-x_i-1}\frac{1}{\beta_i(a)+j}-\sum_{j=0}^{n_i-1}\frac{1}{\alpha_i(a)+\beta_i(a)+j}\right). \tag{38}$$

Then, using the above expressions and the asymptotic property of the ML-estimator $\tilde{a}^0=a^0+\mathbf{1}_m\cdot O_P(1/\sqrt{k})$, $\tilde{b}^0=b^0+\mathbf{1}_m\cdot O_P(1/\sqrt{k})$ completes the proof. □

MLS-estimation of BLM parameters. The partial derivatives of the log-function $l(a,b,X,\varepsilon_0,\varepsilon_1)$ are computed as

$$\frac{\partial l}{\partial a_r}=\sum_{i=1}^{k}\sum_{j=0}^{n_i}\frac{w^i_{x_i j}(\varepsilon_0,\varepsilon_1)\cdot\partial P^i_j(a,b)/\partial a_r}{\sum_{t=0}^{n_i} w^i_{x_i t}(\varepsilon_0,\varepsilon_1)\cdot P^i_t(a,b)},\quad \frac{\partial l}{\partial b_r}=\sum_{i=1}^{k}\sum_{j=0}^{n_i}\frac{w^i_{x_i j}(\varepsilon_0,\varepsilon_1)\cdot\partial P^i_j(a,b)/\partial b_r}{\sum_{t=0}^{n_i} w^i_{x_i t}(\varepsilon_0,\varepsilon_1)\cdot P^i_t(a,b)},$$

$$\frac{\partial l}{\partial \varepsilon_0}=\sum_{i=1}^{k}\sum_{j=0}^{n_i}\frac{\partial w^i_{x_i j}(\varepsilon_0,\varepsilon_1)/\partial\varepsilon_0\cdot P^i_j(a,b)}{\sum_{t=0}^{n_i} w^i_{x_i t}(\varepsilon_0,\varepsilon_1)\cdot P^i_t(a,b)},\quad \frac{\partial l}{\partial \varepsilon_1}=\sum_{i=1}^{k}\sum_{j=0}^{n_i}\frac{\partial w^i_{x_i j}(\varepsilon_0,\varepsilon_1)/\partial\varepsilon_0\cdot P^i_j(a,b)}{\sum_{t=0}^{n_i} w^i_{x_i t}(\varepsilon_0,\varepsilon_1)\cdot P^i_t(a,b)},$$

where

$$\frac{\partial P^i_j(a,b)}{\partial a_r}=P^i_j(a,b)Z_{ir}\,\alpha_i(a)\left(\sum_{l=0}^{j-1}\frac{1}{\alpha_i(a)+l}-\sum_{l=0}^{n_i-1}\frac{1}{\alpha_i(a)+\beta_i(a)+l}\right),$$

$$\frac{\partial P^i_j(a,b)}{\partial b_r}=P^i_j(a,b)Z_{ir}\beta_i(a)\left(\sum_{l=0}^{n_i-j-1}\frac{1}{\beta_i(a)+l}-\sum_{l=0}^{n_i-1}\frac{1}{\alpha_i(a)+\beta_i(a)+l}\right),$$

and $\partial w^i/\partial\varepsilon_0$, $\partial w^i/\partial\varepsilon_1$ are defined above (see the MLS-estimator for the BBM).

Proof of Theorem 8. Under the distortions, the mean square error of forecasting for the classical Bayes predictor can be expressed as

$$\tilde{r}_i^2 = \mathbf{E}\{(p_i - (\alpha_i^0 + x)/(\alpha_i^0 + \beta_i^0 + n_i))^2\},$$

where p_i is the beta random variable with the parameters α_i^0, β_i^0, and the variable x (the distorted sum of binary responses) follows the DBBD with the parameters $n_i, \alpha_i^0, \beta_i^0, \varepsilon_0, \varepsilon_1$. Simplifying the latter expression leads to

$$\tilde{r}_i^2 = \mathbf{E}\{p_i^2\} - 2\frac{\alpha_i^0\mathbf{E}\{p_i\} + \mathbf{E}\{xp_i\}}{\alpha_i^0 + \beta_i^0 + n_i} + \frac{\alpha_i^{0^2} + 2\alpha_i^0\mathbf{E}\{x\} + \mathbf{E}\{x^2\}}{(\alpha_i^0 + \beta_i^0 + n_i)^2}, \quad (39)$$

where $\mathbf{E}\{p_i\} = \alpha_i^0/(\alpha_i^0 + \beta_i^0)$, $\mathbf{E}\{p_i^2\} = \alpha_i^0(\alpha_i^0 + 1)/((\alpha_i^0 + \beta_i^0)(\alpha_i^0 + \beta_i^0 + 1))$, and the mathematical expectations of the random variables x, xp_i, x^2 are

$$\mathbf{E}\{x\} = n\varepsilon_0 + n(1-\varepsilon_0-\varepsilon_1)\cdot\mathbf{E}\{p_i\}, \qquad \mathbf{E}\{xp_i\} = n\varepsilon_0\cdot\mathbf{E}\{p_i\} + n(1-\varepsilon_0-\varepsilon_1)\cdot\mathbf{E}\{p_i^2\}, \quad (40)$$

$$\mathbf{E}\{x^2\} = \mathbf{E}\{x\} + n(n-1)\left(\varepsilon_0^2 + 2\varepsilon_0(1-\varepsilon_0-\varepsilon_1)\cdot\mathbf{E}\{p_i\} + (1-\varepsilon_0-\varepsilon_1)\cdot\mathbf{E}\{p_i^2\}\right). \quad (41)$$

Substituting these formulas to (39) and simplifying the corresponding expression proves the theorem. □

Proof of Theorem 9. Following the proof of Theorem 8, the mean square error of forecasting for the classical Bayes predictor under the distortions (when using the estimates $\hat{\alpha}_i, \hat{\beta}_i$) can be expressed as

$$\tilde{r}_i^2 = \mathbf{E}\{p_i^2\} - 2\frac{\hat{\alpha}_i\mathbf{E}\{p_i\} + \mathbf{E}\{xp_i\}}{\hat{\alpha}_i + \hat{\beta}_i + n_i} + \frac{\hat{\alpha}_i^2 + 2\hat{\alpha}_i\mathbf{E}\{x\} + \mathbf{E}\{x^2\}}{(\hat{\alpha}_i + \hat{\beta}_i + n_i)^2}, \quad (42)$$

where the mathematical expectations $\mathbf{E}\{x\}, \mathbf{E}\{xp_i\}, \mathbf{E}\{x^2\}$ are defined by expressions (40), (41). Then, collecting the coefficients of $\varepsilon_0, \varepsilon_1$ and ε_0^2, $\varepsilon_0\varepsilon_1$, ε_1^2 in expression (42) taking into account the notation (22), (23) proves the theorem. □

Proof of Theorem 10. Using the Bayes formula and Theorem 1, the posterior p.d.f. of the random variable p_i is expressed as:

$$f_{p_i}(x|s,\varepsilon_0,\varepsilon_1) = \frac{\sum_{r=0}^{n_i} w_{sr}^i(\varepsilon_0,\varepsilon_1)\cdot C_{n_i}^r x^r(1-x)^{(n_i-r)}\cdot B(\alpha_i^0,\beta_i^0)^{-1}x^{\alpha_i^0-1}(1-x)^{\beta_i^0-1}}{\int_0^1 \sum_{r=0}^{n_i} w_{sr}^i(\varepsilon_0,\varepsilon_1)\cdot C_{n_i}^r y^r(1-x)^{(n_i-r)}\cdot B(\alpha_i^0,\beta_i^0)^{-1}y^{\alpha_i^0-1}(1-y)^{\beta_i^0-1}dy}.$$

Simplifying this formula using the properties of the beta distribution (Johnson *et al.* 1996) leads to the expression for the forecast p.d.f. (27). Then, calculating the mean of this distribution taking into account the properties of the DBBD gives the predictor

(24). The mean square error of forecasting (26) is derived using the technique given in the proof of Theorem 8 for the obtained predictor. □

8 References

Agresti, A., Booth, J. G., Hobert, J. P. and Caffo, B. (2000). Random effects modeling of categorical response data. *Sociological Methodology*, 30, 27-80.

Brooks, S. P. (2001). On Bayesian analysis and finite mixtures for proportions. *Statistics and Computing*, 11, 179-190.

Collet, D. (2002). *Modeling Binary Data*. London: Champton and Hall/CRC.

Coull, B. A. and Agresti, A. (2000). Random effects modeling of multiple binomial responses using the multivariate binomial logit-normal distribution. *Biometrics*, 56, 73-80.

Copas, J. B. (1988). Binary regression models for contaminated data. *Journal of Royal Statistical Society*, 50B, 225-265.

Danaher, P. J. (1992). Some statistical modeling problems in the advertising industry: a look at media exposure distributions. *The American Statistician*, 46, 241-253.

Demidovich B. P. and Maron, I. A. (1970). *Basics of Computational Mathematics*. Moscow (in Russian).

Diggle, P. J., Heagerty, P., Liang, K.-Y. and Zeger, S. L. (2002). *Analysis of Longitudinal Data*. Oxford : University Press.

Gaba, A. and Winkler, R. L. (1992). Implication of errors in survey data: a Bayesian model. *Management Science.*, 38 (7), 913-925.

Gill, P. S. (2001). A robust mixed linear model analysis for longitudinal data. *Statistics in Medicine*, 19, 975-987.

Hampel, F. R., Rousseeuw, P. J., Ronchetti, E. M. and Stahel, W. A. (1986). *Robust Statistics*. New York: John Wiley and Sons.

Heckman, J. J. and Willis, R. J. (1977). A beta logistic model for the analysis of sequential labor force participation by married women. *Journal of Political Economy*, 85 (11), 27-58.

Huber, P. J. (1981). *Robust Statistics*. New York: Wiley.

Ivchenko, G. I. and Medvedev, U. I. (1984). *Mathematical Statistics*. Moscow (in Russian).

Johnson, N. L., Kotz, S. and Kemp, A. W. (1996). *Univariate Discrete Distributions*. Wiley-Interscience, New York.

Kharin, Yu. (1996). *Robustness in Statistical Pattern Recognition*. Kluwer Academic Publishers, Dordrecht.

Kordzakhia, N., Mishra, G. D. and Reiersolmoen, L. (2001). Robust estimation in the logistic regression model. *Journal of Statistical Planning and Inference*, 98, 211-223.

Liang, K.-Y. and Zeger, S. L. (1986). Longitudinal data analysis using generalized linear models. *Biometrika*, 73, 13-22.

Nathan, D. (1999). A beta-logistic model of presidential influence on voting on civil rights issues in the house of representatives, 1960-1988. *Midwest Annual Meeting Methodology Panel Working Papers*, 32-37.

Neuhaus, J. M. (1999). Bias and efficiency loss due to misclassified responses in binary regression. *Biometrika*, 86, 843-855.

Neuhaus, J. M. (2002). Analysis of clustered and longitudinal binary data subject to response misclassification. *Biometrics*, 58, 675-683.

Pearson, E. S. (1925). Bayes' theorem in the light of experimental sampling. *Biometrika*, 17, 338-442.

Pfeifer, P. E. (1998). On using the beta-logistic model to update response probabilities given nonresponse. *Journal of Interactive Marketing*, 12 (2), 23-32.

Prentice, R. L. (1988). Correlated binary regression with covariates specific to each binary observation. *Biometrics*, 44, 1033-1048.

Prentice, R. L. (1986). Binary Regression Using an Extended Beta-Binomial Distribution, With Discussion of Correlation Induced by Covariate Measurement Errors. *Journal of the American Statistical Association*, 81, 321-327.

Ruckstuhl, A. F. and Welsh, A. H. (2001). Robust fitting of the binomial model. *The Annals of Statistics*, 29, 1117-1136.

Sissors, J. and Lincoln, B. (1994). *Advertising Media Planning*. NTC Business Books. Lincolnwood.

Skellam, J. G. (1948). A probability distribution derived from the binomial distribution by regarding the probability of success as a variable between the sets of trials. *Journal of the Royal Statistical Society*, 10B, 257-261.

Slaton, T. L., Piegorsch, W. W. and Durham, S. D. (2000). Estimation and testing with overdispersed proportions using the beta-logistic regression model of Heckman and Willis. *Biometrics*, 56(1), 125-133.

Swartz, T., Haitovsku, Y., Vexler, A. and Yang, T. Bayesian identifiability and misclassification in multinomial data. *The Canadian Journal of Statistics*, 32, 2004, to appear.

Tripathi, R. C., Gupta, R. C. and Gurland, J. (1994). Estimation of parameters in the beta binomial model. *Ann. Inst. Statist. Math.*, 46, 317-331.

White, H. (1982). Maximum Likelihood Estimation of Misspecified Model. *Econometrica*, 50(1), 1-26.

Zeger, S. L. and Karim, M. R. (1991). Generalized linear models with random effects: A Gibbs sampling approach. *Journal of the American Statistical Association*, 86, 79-86.

SORT 28 (2) July-December 2004, 161-176

Extremes of periodic moving averages of random variables with regularly varying tail probabilities

A. P. Martins, H. Ferreira*

Universidade da Beira Interior

Abstract

We define a family of local mixing conditions that enable the computation of the extremal index of periodic sequences from the joint distributions of k consecutive variables of the sequence. By applying results, under local and global mixing conditions, to the $(2m-1)$–dependent periodic sequence $X_n^{(m)} = \sum_{j=-m}^{m-1} c_j Z_{n-j}$, $n \geq 1$, we compute the extremal index of the periodic moving average sequence $X_n = \sum_{j=-\infty}^{\infty} c_j Z_{n-j}$, $n \geq 1$, of random variables with regularly varying tail probabilities.
This paper generalizes the theory for extremes of stationary moving averages with regularly varying tail probabilities.

MSC: 60G70

Keywords: Periodic moving average processes, extremal index, mixing condition

1 Introduction

The moving average process of the form

$$X_n = \sum_{j=-\infty}^{\infty} c_j Z_{n-j}, \quad n \geq 1, \tag{1.1}$$

with iid real-valued innovations or noise variables $(Z_j)_{j\in\mathbb{Z}}$, includes the popular $ARMA(p,q)$ and $AR(p)$ processes considered in classical time series analysis. Studies of the extreme value behaviour of such processes have been carried out, among others, by Cline (1983), Davis and Resnick (1985, 1988) and Chernick *et al.* (1991).

* *Address for correspondence:* Ana Paula Martins, amartins@noe.ubi.pt. and
Helena Ferreira, ferreira@fenix2.ubi.pt. Departamento de Matemática (Centro de Matemática), Universidade da Beira Interior., 6200 Covilhã, Portugal.
Received: October 2003
Accepted: April 2004

In this paper we are concerned with moving average processes of the form (1.1) but with $\mathbf{Z} = \{Z_j\}_{j\in\mathbb{Z}}$ a $T-$periodic sequence of independent real-valued variables, such that $\overline{F}_i(x) = P(|Z_i| > x)$, $i = 1,\ldots,T$, are regularly varying with exponent $-\alpha$, i.e.,

$$\overline{F}_i(x) = x^{-\alpha}L_i(x), \quad x > 0, \quad i = 1,\ldots,T, \tag{1.2}$$

for some $\alpha > 0$ and $L_i : \mathbb{R}_+ \to \mathbb{R}_+$ slowly varying functions. We also assume the tail balance conditions

$$\lim_{x\to\infty}\frac{P(Z_i > x)}{\overline{F}_i(x)} = p_i, \quad \lim_{x\to\infty}\frac{P(Z_i < -x)}{\overline{F}_i(x)} = q_i, \quad i = 1,\ldots,T, \tag{1.3}$$

for some p_i and $q_i \in [0,1]$ such that $p_i + q_i = 1$, $i = 1,\ldots,T$, and tail equivalence in the following way

$$\lim_{x\to\infty}\frac{P(Z_i > x)}{P(Z_j > x)} = \gamma_{i,j}^{(+)} > 0, \quad \lim_{x\to\infty}\frac{P(Z_i < -x)}{P(Z_j < -x)} = \gamma_{i,j}^{(-)} > 0, \quad i,j = 1,\ldots,T. \tag{1.4}$$

The sequence of real constants $\mathbf{c} = \{c_j\}_{j\in\mathbb{Z}}$ will be taken to satisfy

$$\sum_{j=-\infty}^{\infty} |c_j|^{\delta} < \infty, \tag{1.5}$$

for some $\delta < \min\{\alpha, 1\}$, in order to guarantee the a.s. convergence of (1.1). Notice that conditions (1.3) and (1.4) imply the existence of $\gamma_{i,j} = \lim_{x\to\infty}\frac{P(|Z_i|>x)}{P(|Z_j|>x)}$, $i,j = 1,\ldots,T$.

Extreme value theory known for periodic sequences can then be applied to this moving average sequence $\mathbf{X} = \{X_n\}_{n\geq 1}$, since it is also a $T-$periodic sequence. Alpuim (1988) showed that under Leadbetter's global mixing condition D, the only possible limit laws for the normalized maxima of a $T-$periodic sequence are the three extreme value distributions. Under local mixing conditions $D_T^{(k)}$, $k = 1,2$, Ferreira (1994) studied the extremal behaviour of periodic sequences, and under the weaker local mixing conditions $D_T^{(k)}$, $k \geq 3$, Ferreira and Martins (2003) obtained the expression for the extremal index of a $T-$periodic sequence from the joint distribution of k consecutive variables of the sequence.

We say that for a fixed integer $k \geq 1$ and a sequence of real constants $\mathbf{u} = \{u_n\}_{n\geq 1}$ the condition $D_T^{(k)}(u_n)$ holds for a $T-$periodic sequence $\mathbf{X}$ satisfying Leadbetter's condition $D(u_n)$ (see Leadbetter *et al.* (1983)) with mixing coefficients $\beta_{n,l}$, when there exists a sequence of integers $\mathbf{k} = \{k_n\}_{n\geq 1}$ such that $\lim_{n\to\infty} k_n = \infty$, $\lim_{n\to\infty} k_n\frac{l_n}{n} = 0$, $\lim_{n\to\infty} k_n\beta_{n,l_n} = 0$, and

$$\lim_{n\to\infty} n\frac{1}{T}\sum_{i=1}^{T}\sum_{j=i+k}^{\left[\frac{n}{k_nT}\right]T} P(X_i > u_n \geq M_{i+1,i+k-1}, X_j > u_n) = 0, \tag{1.6}$$

where $M_{i,j} = \max\{X_i, X_{i+1},\ldots,X_j\}$ and $M_{i,j} = -\infty$ for $i > j$.

Under this local dependence condition the extremal index of **X**,

$$\theta_{\mathbf{X}} = \frac{-\log(\lim_{n\to\infty} P(\max_{1\le i\le n} X_i \le u_n))}{\tau},$$

where

$$\tau = \lim_{n\to\infty} n\frac{1}{T}\sum_{i=1}^{T} P(X_i > u_n), \tag{1.7}$$

can be computed from

$$\theta_{\mathbf{X}} = \lim_{n\to\infty} \frac{n\frac{1}{T}\sum_{i=1}^{T} P(X_i > u_n \ge M_{i+1,i+k-1})}{\tau}. \tag{1.8}$$

A sequence **u** satisfying (1.7) is usually denoted by $\mathbf{u}^{(\tau)} = \{u_n^{(\tau)}\}$ for **X**, and its elements are called normalized levels for **X**.

Observe that, when $k \ge 2$, condition (1.6) is implied by

$$\lim_{n\to\infty} S^{(k)}_{\left[\frac{n}{k_n T}\right]} = \lim_{n\to\infty} n\frac{1}{T}\sum_{i=1}^{T}\sum_{j=i+k}^{\left[\frac{n}{k_n T}\right]T} P(X_i > u_n, X_{j-1} \le u_n < X_j) = 0,$$

which limits the distance between exceedances of level u_n, that is, in each interval there can only be more than one exceedance of u_n if separated by less than $k-1$ non-exceedances of u_n. Consequently, the local dependence condition $D_T^{(k)}$, $k \ge 1$, become weaker as the value of k increases.

Our aim in this paper is to use the previous results, that generalize the ones obtained by Chernick *et al.* (1991) for stationary sequences, to obtain the expression for the extremal index of the T-periodic moving average sequence of random variables with regularly varying tail probabilities **X** defined by (1.1) and satisfying certain balance and tail equivalence conditions. To attain this we start by characterizing in Section 2 the behaviour of each tail $P(X_i) > x$, $i = 1, \dots, T$ as $x \to \infty$, and by obtaining sufficient conditions that allow the application of our results to a finite moving average sequence $\mathbf{X}^{(m)}$ that "approximates" **X** as $m \to \infty$. In Section 3 we present our main result which gives the expression of the extremal index of the T-periodic moving average sequence **X**.

The proofs of all theorems presented are given in the Appendix.

2 First results

The first result we present is a simple modification of a theorem found in Resnick (1987) for the stationary case, but crucial for the characterization of the behaviour of each tail $P(X_i > x)$, $i = 1, \dots, T$, as $x \to \infty$, which we present ahead.

Theorem 2.1 *Let* $\mathbf{Z} = \{Z_n\}_{n\in\mathbb{Z}}$ *be a* T*−periodic sequence of independent random variables satisfying (1.2), (1.3) and (1.4) and* $\mathbf{c} = \{c_j\}_{j\in\mathbb{Z}}$ *a sequence of real constants satisfying (1.5). Then for* $i = 1, \ldots, T$ *when* $x \to \infty$

$$\lim_{x\to\infty} \frac{P\left(\sum_{j=-\infty}^{\infty} |c_j Z_{i-j}| > x\right)}{P(|Z_i| > x)} = p_i \sum_{s=0}^{T-1} \gamma_{i-s,i}^{(+)} \, p_{i-s}^{-1} \sum_{j=-\infty}^{\infty} |c_{jT+s}|^{\alpha}. \quad (2.9)$$

The behaviour of each tail $P(X_i > x)$, $i = 1, \ldots, T$, as $x \to \infty$ is vital for the extremal behaviour of the periodic moving average process $\mathbf{X}$. As Embrechts *et al.* (1997), we prove how every r.v. Z_j, $j \in \mathbb{Z}$, has a contribution to each tail $P(X_i > x)$, $i = 1, \ldots, T$.

Theorem 2.2 *Let* $\mathbf{Z} = \{Z_n\}_{n\in\mathbb{Z}}$ *be a* T*−periodic sequence of independent variables satisfying (1.2), (1.3) and (1.4) and* $\mathbf{c} = \{c_j\}_{j\in\mathbb{Z}}$ *a sequence of real constants satisfying (1.5). Then for* $i = 1, \ldots, T$

$$P(X_i > x) \sim x^{-\alpha} L_i(x) \left\{ p_i \sum_{s=0}^{T-1} \gamma_{i-s,i}^{(+)} \sum_{j=-\infty}^{\infty} \left[c_{jT+s}^{+}\right]^{\alpha} + q_i \sum_{s=0}^{T-1} \gamma_{i-s,i}^{(-)} \sum_{j=-\infty}^{\infty} \left[c_{jT+s}^{-}\right]^{\alpha} \right\},$$

where $c_j^+ = max\{c_j, 0\}$ *and* $c_j^- = max\{-c_j, 0\}$.

As we can see, the contribution of the random variables $\mathbf{Z} = \{Z_n\}_{n\in\mathbb{Z}}$ to each tail depends on the size and sign of the respective weight c_j associated to them.

The computation of the extremal index using expression (1.8) requires the validation of a long range and a local mixing condition, which is often a difficult task when considering some sequences, namely moving average sequences. To overcome this difficulty it's useful to consider in these cases an "approximating" sequence $\mathbf{X}^{(m)} = \{X_n^{(m)}\}_{n\geq 1}$ for a fixed integer m, then apply a Slutsky argument and let $m \to \infty$. We can then use the extremal index of this sequence $\mathbf{X}^{(m)}$ to estimate that of $\mathbf{X}$.

Sufficient conditions, to take into consideration such a sequence $\mathbf{X}^{(m)}$, in the periodic case, can be found in the next result, analogous to the one found in Chernick *et al.* (1991) for the stationary case.

Theorem 2.3 *Suppose* $\mathbf{X}$ *and* $\mathbf{X}^{(m)}$, $m \geq 1$ *are* T*−periodic sequences defined on the same probability space such that for some sequences of constants* $\mathbf{u} = \{u_n\}_{n\geq 1}$ *and* $i = 1, \ldots, T$,

$$\lim_{\epsilon\to 0} \limsup_{n\to\infty} nP\left((1-\epsilon)u_n < X_i \leq (1+\epsilon)u_n\right) = 0, \quad (2.10)$$

$$\lim_{m\to\infty} \limsup_{n\to\infty} nP\left(|X_i - X_i^{(m)}| > \epsilon u_n\right) = 0, \quad \epsilon > 0. \quad (2.11)$$

Then

(i) $\lim_{m\to\infty} \limsup_{n\to\infty} \left|P(M(I_n) \leq u_n) - P(M^{(m)}(I_n) \leq u_n)\right| = 0$, *where the supremum is taken over all index sets* $I_n \subset \{1, \ldots, n\}$.

(ii) If condition $D(\mathbf{u})$ holds for $\mathbf{X}^{(m)}$ for each m, then it holds for $\mathbf{X}$ *as well.*

(iii) $\lim_{m\to\infty}\lim_{n\to\infty} n|P(X_i > u_n \geq M_{i+1,i+k-1}) - P(X_i^{(m)} > u_n \geq M_{i+1,i+k-1}^{(m)})| = 0,\ k \geq 2,$

with $M(I) = \max_{j\in I} X_j$ *and* $M^{(m)}(I) = \max_{j\in I} X_j^{(m)}$, *for* $I \subset \{1,\dots,n\}$.

Remark 1 If (2.10) and (2.11) hold with $\mathbf{u}^{(\tau)}$ and $\mathbf{X}^{(m)}$, $m \geq 1$ has extremal index $\theta_{\mathbf{X}^{(m)}}$, then by Theorem 1.3(i) with $I_n = \{1,\dots,n\}$, $\mathbf{X}$ has extremal index $\theta_{\mathbf{X}}$ if and only if $\theta_{\mathbf{X}^{(m)}} \to \theta_{\mathbf{X}}$ as $m \to \infty$.

3 Main result

We are now in conditions to state our main theorem which computes the extremal index $\theta_{\mathbf{X}}$ of a periodic sequence $\mathbf{X}$ of moving averages of random variables with regularly varying tail probabilities. For this, we need to consider a sequence of constants $\mathbf{u} = \{u_n\}_{n\in\mathrm{N}}$ satisfying

$$\lim_{n\to\infty} nP(|Z_i| > u_n) = \tau_i \Bigg/ \left\{ p_i \sum_{s=0}^{T-1} \gamma_{i-s,i}^{(+)} \sum_{j=-\infty}^{\infty} \left[c_{jT+s}^{+}\right]^{\alpha} + q_i \sum_{s=0}^{T-1} \gamma_{i-s,i}^{(-)} \sum_{j=-\infty}^{\infty} \left[c_{jT+s}^{-}\right]^{\alpha} \right\}, \quad (3.1)$$

for given $\tau_i > 0$, $i = 1,\dots,T$. Such a sequence exists by the assumption of regular variation of each $\overline{F}_i$, $i = 1,\dots,T$, and implies, by Theorem 2.2, that $nP(X_i > u_n) \to \tau_i$, $x \to \infty$, $i = 1,\dots,T$, therefore $\mathbf{u} = \mathbf{u}^{(\tau)}$ for $\mathbf{X}$ with $\tau = \frac{1}{T}\sum_{i=1}^{T} \tau_i$.

Theorem 3.1 *Let* $\mathbf{X} = \{X_n\}_{n\geq 1}$ *be a T−periodic moving average sequence as defined in (1.1). Then* $\mathbf{X}$ *has extremal index*

$$\theta_{\mathbf{X}} = \frac{\sum_{i=1}^{T} \gamma_{i,1} \left\{ p_i \sum_{s=0}^{T-1} \gamma_{i-s,i}^{(+)} c_s^{+}(\alpha) + q_i \sum_{s=0}^{T-1} \gamma_{i-s,i}^{(-)} c_s^{-}(\alpha) \right\}}{\sum_{i=1}^{T} \gamma_{i,1} \left\{ p_i \sum_{s=0}^{T-1} \gamma_{i-s,i}^{(+)} \sum_{j=-\infty}^{\infty} \left[c_{jT+s}^{(+)}\right]^{\alpha} + q_i \sum_{s=0}^{T-1} \gamma_{i-s,i}^{(-)} \sum_{j=-\infty}^{\infty} \left[c_{jT+s}^{-}\right]^{\alpha} \right\}},$$

where

$$c_s^{+}(\alpha) = \sum_{j=-\infty}^{\infty} \left(\left[c_{jT+s}^{+}\right]^{\alpha} - \max_{r>jT+s} \{c_r^{+}\}^{\alpha} \right)_{+}$$

and

$$c_s^{-}(\alpha) = \sum_{j=-\infty}^{\infty} \left(\left[c_{jT+s}^{-}\right]^{\alpha} - \max_{r>jT+s} \{c_r^{-}\}^{\alpha} \right)_{+}.$$

This result shows how the balance and tail equivalence parameters influence the value of the mean number of clustered exceedances in these processes.

4 Appendix

Proof (Theorem 2.1). We begin by showing a weaker result, namely for each $i = 1,\dots,T$,

$$\lim_{x\to\infty} \frac{P(|c_1Z_{i-1}| + |c_2Z_{i-2}| + |c_3Z_{i-3}| > x)}{P(|Z_i| > x)}$$
$$= \gamma_{i-1,i}^{(+)}\, p_{i-1}^{-1}\, p_i|c_1|^\alpha + \gamma_{i-2,i}^{(+)}\, p_{i-2}^{-1}\, p_i|c_2|^\alpha + \gamma_{i-3,i}^{(+)}\, p_{i-3}^{-1}\, p_i|c_3|^\alpha. \tag{4.2}$$

We restrict ourselves to three summands with non-zero c_1, c_2, c_3 to show the method, the general case can be proved analogously by induction. For $\delta \in (0, 1/3)$ and $i = 1,\dots,T$,

$$\begin{aligned}
&\{|c_1Z_{i-1}| + |c_2Z_{i-2}| + |c_3Z_{i-3}| > x\}\\
&\quad\subset \{|c_1Z_{i-1}| > (1-\delta)^2x\} \cup \{|c_2Z_{i-2}| > (1-\delta)^2x\} \cup \{|c_3Z_{i-3}| > (1-\delta)^2x\}\\
&\quad\cup \{|c_1Z_{i-1}| > \delta(1-\delta)x, |c_2Z_{i-2}| > \delta(1-\delta)x\}\\
&\quad\cup \{|c_1Z_{i-1}| > \delta(1-\delta)x, |c_3Z_{i-3}| > \delta(1-\delta)x\}\\
&\quad\cup \{|c_2Z_{i-2}| > \delta(1-\delta)x, |c_3Z_{i-3}| > \delta(1-\delta)x\}\\
&\quad\cup \{|c_1Z_{i-1}| > \delta^2x, |c_2Z_{i-2}| > \delta^2x, |c_3Z_{i-3}| > \delta^2x\}.
\end{aligned}$$

Hence, by conditions (1.2), (1.3) and (1.4)

$$\limsup_{x\to\infty} \frac{P(|c_1Z_{i-1}| + |c_2Z_{i-2}| + |c_3Z_{i-3}| > x)}{P(|Z_i| > x)} \le |c_1|^\alpha(1-\delta)^{-2\alpha}p_{i-1}^{-1}\gamma_{i-1,i}^{(+)}\, p_i$$
$$+|c_2|^\alpha(1-\delta)^{-2\alpha}p_{i-2}^{-1}\gamma_{i-2,i}^{(+)}\, p_i + |c_3|^\alpha(1-\delta)^{-2\alpha}p_{i-3}^{-1}\gamma_{i-3,i}^{(+)}\, p_i. \tag{4.3}$$

Moreover,

$$\begin{aligned}
&\{|c_1Z_{i-1}| + |c_2Z_{i-2}| + |c_3Z_{i-3}| > x\}\\
&\quad\supset \{|c_1Z_{i-1}| > (1+\delta)^2x, |c_2Z_{i-2}| + |c_3Z_{i-3}| \le \delta x\}\\
&\quad\cup \{|c_1Z_{i-1}| \le \delta x, |c_2Z_{i-2}| > (1+\delta)^2x, |c_3Z_{i-3}| \le \delta(1+\delta)x\}\\
&\quad\cup \{|c_1Z_{i-1}| \le \delta x, |c_2Z_{i-2}| \le \delta(1+\delta)x, |c_3Z_{i-3}| > (1+\delta)^2x\}.
\end{aligned}$$

Hence

$$\liminf_{x\to\infty} \frac{P(|c_1Z_{i-1}| + |c_2Z_{i-2}| + |c_3Z_{i-3}| > x)}{P(|Z_i| > x)} \ge |c_1|^\alpha(1+\delta)^{-2\alpha}p_{i-1}^{-1}\gamma_{i-1,i}^{(+)}\, p_i$$
$$+|c_2|^\alpha(1+\delta)^{-2\alpha}p_{i-2}^{-1}\gamma_{i-2,i}^{(+)}\, p_i + |c_3|^\alpha(1+\delta)^{-2\alpha}p_{i-3}^{-1}\gamma_{i-3,i}^{(+)}\, p_i. \tag{4.4}$$

Letting $\delta \to 0$ in (4.3) and (4.4) we obtain (4.2). Notice that in the case $T = 2$ we have $\gamma_{i-1,i}^{(+)} = \gamma_{i-3,i}^{(+)}$, $p_{i-1} = p_{i-3}$, for $i = 1, 2$ and $p_1 = p_3$.

We must leap now from (4.2) to (2.9). For $x > 0$ and $I_1 = \{0, \ldots, T-1\}$, write

$$P\left(\sum_{j=-\infty}^{\infty} |c_j Z_{i-j}| > x\right) = P\left(\sum_{s=0}^{T-1}\sum_{j=-\infty}^{\infty} |c_{jT+s} Z_{i-jT-s}| > x\right)$$

$$= P\left(\sum_{s=0}^{T-1}\sum_{j=-\infty}^{\infty} |c_{jT+s} Z_{i-jT-s}| > x, \max_{s\in I_1}\max_{j\in\mathbb{R}} |c_{jT+s} Z_{i-jT-s}| > x\right)$$

$$+P\left(\sum_{s=0}^{T-1}\sum_{j=-\infty}^{\infty} |c_{jT+s} Z_{i-jT-s}| > x, \max_{s\in I_1}\max_{j\in\mathbb{R}} |c_{jT+s} Z_{i-jT-s}| \leq x\right)$$

$$\leq P\left(\bigcup_{s=0}^{T-1}\bigcup_{j=-\infty}^{\infty} \{|c_{jT+s} Z_{i-jT-s}| > x\}\right)$$

$$+P\left(\sum_{s=0}^{T-1}\sum_{j=-\infty}^{\infty} |c_{jT+s} Z_{i-jT-s}| \mathbb{1}_{\{|c_{jT+s} Z_{i-jT-s}| \leq x\}} > x\right).$$

Applying Markov's inequality to the second term on the right hand side, we obtain for $i = 1, \ldots, T$,

$$P\left(\sum_{j=-\infty}^{\infty} |c_j Z_{i-j}| > x\right)\Big/P(|Z_i| > x)$$

$$\leq \sum_{s=0}^{T-1}\sum_{j=-\infty}^{\infty} P(|Z_{i-s}| > |c_{jT+s}|^{-1}x)\Big/P(|Z_i| > x)$$

$$+\frac{1}{x}\sum_{s=0}^{T-1}\sum_{j=-\infty}^{\infty} |c_{jT+s}| E(|Z_{i-s}| \mathbb{1}_{\{|Z_{i-s}| \leq |c_{jT+s}|^{-1}x\}})\Big/P(|Z_i| > x)$$

$$= I(x) + J(x). \tag{4.5}$$

For $I(x)$, since for $i = 1, \ldots, T$ and $s \in I_1$, $P(|Z_{i-s}| > x) \in RV_{-\alpha}$, we have that for all $s \in I_1$ and $j \in \mathbb{R}$ such that $|c_{jT+s}| < 1$ (i.e., all but a finite number) there exists x_0 such that $x > x_0$ implies

$$\frac{P(|Z_{i-s}| > |c_{jT+s}|^{-1}x)}{P(|Z_i| > x)} \leq (1+\rho)|c_{jT+s}|^{\rho} p_{i-s}^{-1} \gamma_{i-s,i}^{(+)} p_i,$$

for each $i = 1, \ldots, T$. This bound is summable because of (1.5) and hence by dominated convergence

$$\lim_{x\to\infty} I(x) = p_i \sum_{s=0}^{T-1} \gamma_{i-s,i}^{(+)} p_{i-s}^{-1} \sum_{j=-\infty}^{\infty} |c_{jT+s}|^{\alpha}.$$

For $J(x)$ lets start by considering $0 < \alpha < 1$. From an integration by parts

$$\frac{E(|Z_{i-s}|\mathbb{1}_{\{|Z_{i-s}|\le x\}})}{xP(|Z_{i-s}| > x)} = \frac{\int_0^x P(|Z_{i-s}| > u)du}{xP(|Z_{i-s}| > x)} - 1,$$

and since $P(|Z_{i-s}| > x) \in RV_{-\alpha}$ for all $i = 1, \ldots, T$, $s \in I_1$, by applying Karamata's Theorem this converges as $x \to \infty$ to $\alpha(1-\alpha)^{-1}$. Thus $E(|Z_{i-s}|\mathbb{1}_{\{|Z_{i-s}|\le x\}}) \in RV_{1-\alpha}$ and hence we have, for all but a finite number of s and j's, that for x sufficiently large and some constant $K' > 0$,

$$\begin{aligned}|c_{jT+s}|\frac{E(|Z_{i-s}|\mathbb{1}_{\{|Z_{i-s}|\le|c_{jT+s}|^{-1}x\}})}{xP(|Z_i| > x)} &\le K'|c_{jT+s}|(|c_{jT+s}|^{-1})^{1-\alpha+\alpha-\rho}\, p_{i-s}^{-1}\gamma_{i-s,i}^{(+)}\, p_i \\ &= K'|c_{jT+s}|^{\rho}\, p_{i-s}^{-1}\gamma_{i-s,i}^{(+)}\, p_i,\end{aligned}$$

which is summable in s and j. So we conclude

$$\limsup_{x\to\infty} J(x) \le K'\, p_i \sum_{s=0}^{T-1} \gamma_{i-s,i}^{(+)}\, p_{i-s}^{-1} \sum_{j=-\infty}^{\infty} |c_{jT+s}|^{\alpha},$$

and hence with $0 < \alpha < 1$ for some $K' > 0$

$$\limsup_{x\to\infty} \frac{P\left(\sum_{j=-\infty}^{\infty} |c_j Z_{i-j}| > x\right)}{P(|Z_i| > x)} \le (K'+1)p_i \sum_{s=0}^{T-1} \gamma_{i-s,i}^{(+)}\, p_{i-s}^{-1} \sum_{j=-\infty}^{\infty} |c_{jT+s}|^{\alpha}. \tag{4.6}$$

If $\alpha \ge 1$, we get a similar inequality by reduction to the case $0 < \alpha < 1$ as follows: Pick $\beta \in (\alpha, \alpha\delta^{-1})$ and consider $c = \sum_j |c_j|$ and $p_j = |c_j|/c$. By Jensen's inequality we get

$$\left(\sum_{j=-\infty}^{\infty} |c_j Z_{i-j}|\right)^{\beta} = c^{\beta}\left(\sum_{j=-\infty}^{\infty} p_j |Z_{i-j}|\right)^{\beta} \le c^{\beta-1} \sum_{j=-\infty}^{\infty} |c_j|\, |Z_{i-j}|^{\beta}.$$

Then, by (4.5) we can write for $i = 1, \ldots, T$ and $\beta \in (\alpha, \alpha\delta^{-1})$

$$\begin{aligned}&P\left(\sum_{j=-\infty}^{\infty} |c_j Z_{i-j}| > x\right)\Big/P(|Z_i| > x) \\ &\le P\left(\sum_{s=0}^{T-1}\sum_{j=-\infty}^{\infty} |c_{jT+s}|\, |Z_{i-jT-s}|^{\beta} > c^{1-\beta}x^{\beta}\right)\Big/P(|Z_i|^{\beta} > x^{\beta}) \\ &\le \sum_{s=0}^{T-1}\sum_{j=-\infty}^{\infty} P(|Z_{i-s}|^{\beta} > |c_{jT+s}|^{-1}c^{1-\beta}x^{\beta})\big/P(|Z_i|^{\beta} > x^{\beta}) \\ &+ \frac{1}{c^{1-\beta}x^{\beta}} \sum_{s=0}^{T-1}\sum_{j=-\infty}^{\infty} |c_{jT+s}| E(|Z_{i-s}|^{\beta}\,\mathbb{1}_{\{|Z_{i-s}|^{\beta}\le|c_{jT+s}|^{-1}c^{1-\beta}x^{\beta}\}})\big/P(|Z_i|^{\beta} > x^{\beta}).\end{aligned}$$

Now, as before, since $P(|Z_{i-s}|^\beta > x) \in RV_{-\alpha\beta^{-1}}$ with $\delta < \alpha\beta^{-1} < 1$, for all $i = 1, \dots, T$, $s \in I_1$,

$$\limsup_{x\to\infty} \frac{P\left(\sum_{j=-\infty}^{\infty} |c_j Z_{i-j}| > x\right)}{P(|Z_i| > x)} \le (1+K'')p_i \sum_{s=0}^{T-1} \gamma_{i-s,i}^{(+)} p_{i-s}^{-1} \sum_{j=-\infty}^{\infty} |c_{jT+s}|^{\alpha\beta^{-1}} c^{\alpha(1-\beta^{-1})} < \infty, \tag{4.7}$$

for some constant $K'' > 0$, which is similar to (4.6).

We are now in conditions to prove (2.9): For any integer $m = KT$ with $K \ge 1$ we have the obvious extension of (4.2)

$$\frac{P\left(\sum_{j=-\infty}^{\infty} |c_j Z_{i-j}| > x\right)}{P(|Z_i| > x)} \ge \frac{P\left(\sum_{j=-m}^{m-1} |c_j Z_{i-j}| > x\right)}{P(|Z_i| > x)}$$

$$\underset{x\to\infty}{\longrightarrow} \quad p_i \sum_{j=-m}^{m-1} \gamma_{i-j,i}^{(+)} p_{i-j}^{-1} |c_j|^\alpha = p_i \sum_{s=0}^{T-1} \gamma_{i-s,i}^{(+)} p_{i-s}^{-1} \sum_{j=-K}^{K-1} |c_{jT+s}|^\alpha,$$

and since K is arbitrary

$$\liminf_{x\to\infty} \frac{P\left(\sum_{j=-\infty}^{\infty} |c_j Z_{i-j}| > x\right)}{P(|Z_i| > x)} \ge p_i \sum_{s=0}^{T-1} \gamma_{i-s,i}^{(+)} p_{i-s}^{-1} \sum_{j=-\infty}^{\infty} |c_{jT+s}|^\alpha.$$

On the other hand, for any $\epsilon > 0$, $I_2 = \{-m, \dots, m-1\}$ and $I_2^* = \mathbb{N}_0 \backslash I_2$

$$\frac{P\left(\sum_{j=-\infty}^{\infty} |c_j Z_{i-j}| > x\right)}{P(|Z_i| > x)} \le \frac{P\left(\sum_{j\in I_2} |c_j Z_{i-j}| > (1-\epsilon)x\right)}{P(|Z_i| > x)} + \frac{P\left(\sum_{j\in I_2^*} |c_j Z_{i-j}| > \epsilon x\right)}{P(|Z_i| > x)}$$

and so from (4.2) and (4.6) for some $K' > 0$

$$\limsup_{x\to\infty} \frac{P\left(\sum_{j=-\infty}^{\infty} |c_j Z_{i-j}| > x\right)}{P(|Z_i| > x)} \le (1-\epsilon)^{-\alpha} p_i \sum_{s=0}^{T-1} \gamma_{i-s,i}^{(+)} p_{i-s}^{-1} \sum_{j=-K}^{K-1} |c_{jT+s}|^\alpha$$

$$+(K'+1)\epsilon^{-\alpha} p_i \sum_{s=0}^{T-1} \gamma_{i-s,i}^{(+)} p_{i-s}^{-1} \sum_{j\notin\{-K,\dots,K-1\}} |c_{jT+s}|^\alpha,$$

for the case $0 < \alpha < 1$, with a similar bound provided by (4.7) when $\alpha \ge 1$. Let $K \to \infty$ and then send $\epsilon \to 0$ to obtain for $i = 1, \dots, T$

$$\limsup_{x\to\infty} \frac{P\left(\sum_{j=-\infty}^{\infty} |c_j Z_{i-j}| > x\right)}{P(|Z_i| > x)} \le p_i \sum_{s=0}^{T-1} \gamma_{i-s,i}^{(+)} p_{i-s}^{-1} \sum_{j=-\infty}^{\infty} |c_{jT+s}|^\alpha,$$

which combined with the lim inf statement proves (2.9). □

Proof (Theorem 2.2). For $m = KT$ with $K \ge 1$ arbitrary, lets consider the T−periodic sequence $\mathbf{X}^{(m)} = \{X_n^{(m)}\}_{n\ge1}$ of finite moving averages of the form

$$X_n^{(m)} = \sum_{j=-m}^{m-1} c_j Z_{n-j}, \quad X_n^{*(m)} = X_n - X_n^{(m)}. \tag{4.8}$$

For $X_i^{(m)}, i = 1, \ldots, T, m$ defined in this way we have for $\epsilon \in (0,1)$,

$$P(X_i^{(m)} > (1+\epsilon)x) - P\left(\sum_{j \notin \{-m,\ldots,m-1\}} |c_j Z_{i-j}| \geq \epsilon x\right) \tag{4.9}$$

$$\leq P(X_i^{(m)} > (1+\epsilon)x) - P(X_i^{*(m)} \leq -\epsilon x)$$

$$\leq P(X_i^{(m)} > (1+\epsilon)x, X_i^{*(m)} > -\epsilon x)$$

$$\leq P(X_i > x)$$

$$\leq P(X_i^{(m)} > (1-\epsilon)x) + P(X_i^{*(m)} > \epsilon x)$$

$$\leq P(X_i^{(m)} > (1-\epsilon)x) + P\left(\sum_{j \notin \{-m,\ldots,m-1\}} |c_j Z_{i-j}| > \epsilon x\right). \tag{4.10}$$

Theorem 1.1 implies that

$$\lim_{m\to\infty} \lim_{x\to\infty} P\left(\sum_{j \notin \{-m,\ldots,m-1\}} |c_j Z_{i-j}| \geq \epsilon x\right) \Big/ P(|Z_i| > x)$$

$$= \lim_{K\to\infty} p_i \sum_{s=0}^{T-1} \gamma_{i-s,i}^{(+)} p_{i-s}^{-1} \sum_{j \notin \{-K,\ldots,K-1\}} |c_{jT+s}|^\alpha = 0.$$

The latter relation, (4.9) and (4.10) show that it's suffice to prove, for every $m = KT$, $K \geq 1$ that

$$P(X_i^{(m)} > x) \sim x^{-\alpha} L_i(x) \left\{ p_i \sum_{s=0}^{T-1} \gamma_{i-s,i}^{(+)} \sum_{j=-K}^{K-1} \left[c_{jT+s}^+\right]^\alpha + q_i \sum_{s=0}^{T-1} \gamma_{i-s,i}^{(-)} \sum_{j=-K}^{K-1} \left[c_{jT+s}^-\right]^\alpha \right\}.$$

As in the proof of Theorem 1.1, by applying (1.2), (1.3) and (1.4), we have for $\delta \in (0, 1/3)$ and $i = 1, \ldots, T$,

$$\limsup_{x\to\infty} \frac{P(c_1 Z_{i-1} + c_2 Z_{i-2} + c_3 Z_{i-3} > x)}{P(|Z_i| > x)}$$

$$\leq (1-\delta)^{-2\alpha} p_i \sum_{j=1}^{3} \gamma_{i-j,i}^{(+)} \left[c_j^+\right]^\alpha + (1-\delta)^{-2\alpha} q_i \sum_{j=1}^{3} \gamma_{i-j,i}^{(-)} \left[c_j^-\right]^\alpha, \tag{4.11}$$

and

$$\liminf_{x\to\infty} \frac{P(c_1 Z_{i-1} + c_2 Z_{i-2} + c_3 Z_{i-3} > x)}{P(|Z_i| > x)}$$

$$\geq (1+\delta)^{-2\alpha} p_i \sum_{j=1}^{3} \gamma_{i-j,i}^{(+)} \left[c_j^+\right]^\alpha + (1-\delta)^{-2\alpha} q_i \sum_{j=1}^{3} \gamma_{i-j,i}^{(-)} \left[c_j^-\right]^\alpha. \tag{4.12}$$

Letting $\delta \to 0$ in (4.11) and (4.12) concludes the proof. □

Proof (Theorem 2.3). *(i)* For $\epsilon > 0$

$$|P(M(I_n) \le u_n) - P(M^{(m)}(I_n) \le u_n)|$$
$$\le P((1-\epsilon)u_n < M(I_n) \le (1+\epsilon)u_n) + P(|M(I_n) - M^{(m)}(I_n)| > \epsilon u_n)$$
$$\le n\sum_{i=1}^{T} P((1-\epsilon)u_n < X_i \le (1+\epsilon)u_n) + n\sum_{i=1}^{T} P(|X_i - X_i^{(m)}| > \epsilon u_n).$$

Following *(i)* from (2.10) and (2.11).

(ii) Let $\lambda \in (0,1)$, $A \subset \{1,\dots,k\}$ and $B \subset \{k+[n\lambda],\dots,n\}$, $k \le n - [n\lambda]$. Taking the suprema over all A and B we have by the triangle inequality

$$\limsup_{n\to\infty} \{|P(M(A\cup B) \le u_n) - P(M^{(m)}(A\cup B) \le u_n)|\}$$
$$\le \lim_{m\to\infty} \limsup_{n\to\infty} \{|P(M(A\cup B) \le u_n) - P(M^{(m)}(A\cup B) \le u_n)|\}$$
$$+ \lim_{m\to\infty} \limsup_{n\to\infty} \{|P(M^{(m)}(A) \le u_n)P(M^{(m)}(B) \le u_n) - P(M(A) \le u_n)P(M(B) \le u_n)|\}$$
$$+ \lim_{m\to\infty} \limsup_{n\to\infty} \{|P(M^{(m)}(A\cup B) \le u_n) - P(M^{(m)}(A) \le u_n)P(M^{(m)}(B) \le u_n)|\} = o(1),$$

by *(i)* and the fact that $D(\mathbf{u})$ holds $\mathbf{X}^{(m)}$, for all m. Thus, by Lemma 3.2.1 of Leadbetter *et al.* (1983), $D(\mathbf{u})$ holds for $\mathbf{X}$, with $\beta_{n,[n\lambda]} = 0$.

(iii) Since

$$n|P(X_i > u_n \ge M_{i+1,i+k-1}) - P(X_i^{(m)} > u_n \ge M_{i+1,i+k-1}^{(m)})|$$
$$\le n|P(M_{i+1,i+k-1} \le u_n) - P(M_{i+1,i+k-1}^{(m)} \le u_n)| + n|P(M_{i,i+k-1} \le u_n) - P(M_{i,i+k-1}^{(m)} \le u_n)|,$$

(iii) follows immediately from *(i)*. □

Proof (Theorem 3.1). Lets consider again the T-periodic sequence $\mathbf{X}^{(m)}, m \ge 1$ of finite moving averages defined in (4.8). Since $\mathbf{X}^{(m)}$ is $(2m-1)$-dependent it verifies $D(\mathbf{u})$ with mixing coefficient $\beta_{n,l_n} = 0$, for $l_n \ge 2m$.

From (3.1) and Theorem 1.2, it follows that

$$nP(X_i > u_n) \underset{n\to\infty}{\longrightarrow} \tau_i, \quad i = 1,\dots,T.$$

Hence, by considering $c_j = 0$ for $j \notin I_2 = \{-m,\dots,m-1\}$, we also have

$$nP(X_i^{(m)} > u_n) \underset{n\to\infty}{\longrightarrow} \tau_i, \quad i = 1,\dots,T.$$

In this way, $D_T^{(2m)}(\mathbf{u})$ also holds for $\mathbf{X}^{(m)}$ since for $\mathbf{k} = \{k_n\}_{n\in\mathbb{N}}$ as in the definition of $D_T^{(2m)}$ with $l_n \equiv 2m$ we have for $r_n' = \left[\frac{n}{k_n T}\right]$

$$S_{r_n'}^{(2m)} = n\frac{1}{T}\sum_{i=1}^{T}\sum_{j=i+2m}^{r_n'T} P(X_i^{(m)} > u_n, X_{j-1}^{(m)} \le u_n < X_j^{(m)})$$
$$\le r_n' n\frac{1}{T}\sum_{i=1}^{T} P(X_i^{(m)} > u_n)P(X_{i+2m}^{(m)} > u_n) = o(1).$$

We can then use (1.8) to compute the extremal index of $\mathbf{X}^{(m)}$.

For $i = 1, \dots, T$,

$$P(X_i^{(m)} > u_n \ge M_{i+1,i+2m-1}^{(m)} = P(M_{i+1,i+2m-1}^{(m)} \le u_n, \max_{j\in I_2} \{c_jZ_{i-j}\} > u_n)$$
$$- P(M_{i,i+2m-1}^{(m)} \le u_n, \max_{j\in I_2} \{c_jZ_{i-j}\} > u_n)$$
$$+ P(X_i^{(m)} > u_n \ge M_{i+1,i+2m-1}^{(m)}, \max_{j\in I_2} \{c_jZ_{i-j}\} \le u_n). \quad (4.13)$$

Let us first note that for any $\epsilon > 0$,

$$P(X_i^{(m)} > u_n \ge M_{i+1,i+2m-1}^{(m)}, \max_{j\in I_2} \{c_jZ_{i-j}\} \le u_n)$$
$$\le P(X_i^{(m)} > u_n, \max_{j\in I_2} \{c_jZ_{i-j}\} \le (1-\epsilon)u_n) + P((1-\epsilon)u_n < \max_{j\in I_2} \{c_jZ_{i-j}\} \le u_n),$$

and

$$P(X_i^{(m)} > u_n, \max_{j\in I_2} \{c_jZ_{i-j}\} \le (1-\epsilon)u_n)$$
$$= P\left(\bigcup_{s=-m}^{m-1} \{X_i^{(m)} > u_n, c_sZ_{i-s} = \max_{j\in I_2} c_jZ_{i-j} \le (1-\epsilon)u_n\}\right)$$
$$\le \sum_{s=-m}^{m-1} P(X_i^{(m)} > u_n, c_sZ_{i-s} \le (1-\epsilon)u_n)$$
$$\le \sum_{s=-m}^{m-1} P\left(\sum_{\substack{k=-m\\k\ne s}}^{m-1} \min\{c_kZ_{i-k}, c_sZ_{i-s}\} > \epsilon u_n\right)$$
$$\le \sum_{s=-m}^{m-1} P\left(\sum_{\substack{k=-m\\k\ne s}}^{m-1} c_kZ_{i-k} > \epsilon u_n, \sum_{\substack{k=-m\\k\ne s}}^{m-1} c_sZ_{i-s} > \epsilon u_n\right)$$
$$\le \sum_{s=-m}^{m-1}\sum_{\substack{k=-m\\k\ne s}}^{m-1} P\left(c_kZ_{i-k} > \epsilon u_n(2m-1)^{-1}\right) P\left(c_sZ_{i-s} > \epsilon u_n(2m-1)^{-1}\right) = O(n^{-2}).$$
$$(4.14)$$

On the other hand, for $\rho > 0$

$$\lim_{n\to\infty} nP\left(\max_{j\in I_2} c_jZ_{i-j} > \rho u_n\right) = \lim_{n\to\infty} n\left\{\sum_{j\in S_+} P(Z_{i-j} > \rho c_j^{-1}u_n) + \sum_{j\in S_-} P(Z_{i-j} < \rho c_j^{-1}u_n)\right\}$$

$$= \frac{\tau_i p_i \rho^{-\alpha}\sum_{j\in S_+} c_j^\alpha \gamma_{i-j,i}^{(+)} + \tau_i q_i \rho^{-\alpha}\sum_{j\in S_-}(-c_j)^\alpha \gamma_{i-j,i}^{(-)}}{p_i\sum_{s=0}^{T-1}\gamma_{i-s,i}^{(+)}\sum_{j=-\infty}^{\infty}\left[c_{jT+s}^+\right]^\alpha + q_i\sum_{s=0}^{T-1}\gamma_{i-s,i}^{(-)}\sum_{j=\infty}^{\infty}\left[c_{jT+s}^-\right]^\alpha}$$

where $S_+ = \{j : c_j \geq 0, j \in I_2\}$ and $S_- = \{j : c_j < 0, j \in I_2\}$.
Hence,

$$\lim_{\epsilon\to 0}\lim_{n\to\infty} nP\left((1-\epsilon)u_n < \max_{j\in I_2} c_jZ_{i-j} \leq u_n\right)$$

$$= \lim_{\epsilon\to 0}\lim_{n\to\infty} nP\left(\max_{j\in I_2} c_jZ_{i-j} > (1-\epsilon)u_n\right) - \lim_{n\to\infty} nP\left(\max_{j\in I_2} c_jZ_{i-j} > u_n\right) = 0, \quad (4.15)$$

and so from (4.14) and (4.15) it follows that for $i = 1, \ldots, T$,

$$\lim_{n\to\infty} nP\left(X_i^{(m)} > u_n \geq M_{i+1,i+2m-1},\ \max_{j\in I_2} c_jZ_{i-j} \leq u_n\right) = 0. \quad (4.16)$$

By a similar analysis we deduce that

$$\lim_{n\to\infty} nP\left(M_{i,i+2m-1} \leq u_n,\ \max_{j\in I_2} c_jZ_{i-j} > u_n\right) = 0. \quad (4.17)$$

Combining (4.13), (4.16) and (4.17) we obtain

$$\lim_{n\to\infty} nP\left(X_i^{(m)} > u_n \geq M_{i+1,i+2m-1}^{(m)}\right) \quad (4.18)$$

$$= \lim_{n\to\infty} nP\left(M_{i+1,i+2m-1}^{(m)} \leq u_n,\ \max_{j\in I_2} c_jZ_{i-j} > u_n\right)$$

$$= \lim_{n\to\infty} n\left\{\sum_{j\in S_+} P\left(c_jZ_{i-j} > u_n \geq M_{i+1,i+2m-1}^{(m)}\right) + \sum_{j\in S_-} P\left(c_jZ_{i-j} > u_n \geq M_{i+1,i+2m-1}^{(m)}\right)\right\}.$$

If $j \in S_+$ and $c_jZ_{i-j} > u_n$ for each $i = 1, \ldots, T$, then the condition $M_{i+1,i+2m-1} \leq u_n$ is essentially $\max(0, \max_{i+1\leq s\leq i+2m-1} c_{s-i+j})Z_{i-j} = c_j^+(2m)Z_{i-j}$. In the same way, if $j \in S_-$ and $c_jZ_{ij} > u_n$ for each $i = 1, \ldots, T$, then the condition $M_{i+1,i+2m-1} \leq u_n$ can be replaced by $\max(0, \max_{i+1\leq s\leq i+2m-1}(-c_{s-i+j})(-Z_{i-j})) = -c_j^-(2m)Z_{i-j}$. Using these arguments it is straightforward that (4.18) equals

$$= \lim_{n\to\infty} n \left\{ \sum_{j\in S_+} P(c_j Z_{i-j} > u_n \geq c_j^+(2m) Z_{i-j}) + \sum_{j\in S_-} P(c_j Z_{i-j} > u_n \geq (-c_j^-(2m)) Z_{i-j}) \right\}$$

$$= \frac{\tau_i \left\{ p_i \sum_{j=-m}^{m-1} \gamma_{i-j,i}^{(+)}([c_j^+]^\alpha - [c_j^+(2m)]^\alpha)_+ + q_i \sum_{j=-m}^{m-1} \gamma_{i-j,i}^{(-)}([c_j^-]^\alpha - [c_j^-(2m)]^\alpha)_+ \right\}}{p_i \sum_{s=0}^{T-1} \gamma_{i-s,i}^{(+)} \sum_{j=-\infty}^{\infty} [c_{jT+s}^+]^\alpha + q_i \sum_{s=0}^{T-1} \gamma_{i-s,i}^{(-)} \sum_{j=\infty}^{\infty} [c_{jT+s}^-]^\alpha}$$

where in the definition of $c_j^\pm(2m)$ we use the convention $c_j = 0$ for $j \notin I_2 = \{-m, \ldots, m-1\}$, with in particular, $c_{m-1}^\pm(2m) = 0$.

From this, it follows immediately the subsequent expression for the extremal index of $\mathbf{X}^{(m)}$

$$\theta_{\mathbf{X}}^{(m)} = \frac{\sum_{i=1}^{T} \dfrac{\tau_i \left\{ p_i \sum_{j=-m}^{m-1} \gamma_{i-j,i}^{(+)}([c_j^+]^\alpha - [c_j^+(2m)]^\alpha)_+ + q_i \sum_{j=-m}^{m-1} \gamma_{i-j,i}^{(-)}([c_j^-]^\alpha - [c_j^-(2m)]^\alpha)_+ \right\}}{p_i \sum_{s=0}^{T-1} \gamma_{i-s,i}^{(+)} \sum_{j=-\infty}^{\infty} [c_{jT+s}^+]^\alpha + q_i \sum_{s=0}^{T-1} \gamma_{i-s,i}^{(-)} \sum_{j=-\infty}^{\infty} [c_{jT+s}^-]^\alpha}}{\sum_{i=1}^{T} \tau_i} \tag{4.19}$$

Now considering

$$A_i = p_i \sum_{s=0}^{T-1} \gamma_{i-s,i}^{(+)} \sum_{j=-\infty}^{\infty} [c_{jT+s}^+]^\alpha + q_i \sum_{s=0}^{T-1} \gamma_{i-s,i}^{(-)} \sum_{j=-\infty}^{\infty} [c_{jT+s}^-]^\alpha, \quad i = 1, \ldots, T,$$

by (3.1) we can establish the following relation

$$\frac{\tau_i}{A_i} = \frac{\tau_j}{A_j} \gamma_{i,j}, \quad i, j = 1, \ldots, T.$$

Using this relation in (4.19) we obtain the next simplified expression for the extremal index of $\mathbf{X}^{(m)}$

$$\theta_{\mathbf{X}}^{(m)} = \frac{\sum_{i=1}^{T} \left\{ \gamma_{i,1}\, p_i \sum_{s=0}^{T-1} \gamma_{i-s,i}^{(+)} c_s^{(K)}(\alpha)^+ + \gamma_{i,1}\, q_i \sum_{s=0}^{T-1} \gamma_{i-s,i}^{(-)} c_s^{(K)}(\alpha)^- \right\}}{\sum_{i=1}^{T} \left\{ \gamma_{i,1}\, p_i \sum_{s=0}^{T-1} \gamma_{i-s,i}^{(+)} \sum_{j=-\infty}^{\infty} [c_{jT+s}^{(+)}]^\alpha + \gamma_{i,1}\, q_i \sum_{s=0}^{T-1} \gamma_{i-s,i}^{(-)} \sum_{j=-\infty}^{\infty} [c_{jT+s}^-]^\alpha \right\}}$$

where $c_s^{(K)}(\alpha)^+ = \sum_{j=-K}^{K-1} ([c_{jT+s}^+]^\alpha - [c_{jT+s}^+(2m)]^\alpha)_+$ and $c_s^{(K)}(\alpha)^- = \sum_{j=-K}^{K-1} ([c_{jT+s}^-]^\alpha - [c_{jT+s}^-(2m)]^\alpha)_+$, $s = 0, \ldots, T-1$.

It follows by an easy check that $\theta_{\mathbf{X}^{(m)}} \to \theta_{\mathbf{X}}$ as $m = KT \to \infty$, hence by Theorem 1.3 and the remark immediately following it we obtain the result upon showing (2.10) and (2.11) which is straightforward. □

Acknowledgments We are grateful to the referee for the overall report which helped in improving the final form of this paper.

References

Alpuim, M. T. (1988). *Contribuções à teoria de extremos em sucessões dependentes*. Ph. D. Thesis. DEIOC. University of Lisbon.

Chernick, M. R., Hsing, T. and McCormick, W. (1991). Calculating the extremal index for a class of stationary sequences. *Adv. Appl. Prob.*, 23, 845-850.

Cline, D. (1983). Infinite series of random variables with regularly varying tails. *Tech. Report of Univ. of British Columbia*, 83-24.

Davis, R. and Resnick, S. I. (1985). Limite theory for moving averages of random variables with regularly varying tail probabilities. *Ann. Prob.*, 13, 179-195.

Embrechts, P., Klüppelberg, C. and Mikosch, T. (1997). *Modelling Extremal Events*. Springer-Verlag.

Ferreira, H. (1994). Multivariate extreme values in T-periodic random sequences under mild oscilation restrictions. *Stochastic Process. Appl.*, 14, 111-125.

Ferreira, H. and Martins, A. P. (2003). The extremal index of sub-sampled periodic sequences with strong local dependence. *Revstat*, 1, 15-24.

Leadbetter, M. R., Lindgren, G. and Rootzén, H. (1983). *Extremes and Related Properties of Random Sequences and Processes*. New York: Springer-Verlag.

Resnick, S. I. (1987). *Extreme Values, Regular Variation and Point Processes*. New York: Springer-Verlag.

SORT 28 (2) July-December 2004, 177-190

Asymptotically optimal filtering in linear systems with fractional Brownian noises⋆

M. L. Kleptsyna*, A. Le Breton**, M. Viot***

* *Université du Maine,* **,*** *Université J. Fourier.*

Abstract

In this paper, the filtering problem is revisited in the basic Gaussian homogeneous linear system driven by fractional Brownian motions. We exhibit a simple approximate filter which is asymptotically optimal in the sense that, when the observation time tends to infinity, the variance of the corresponding filtering error converges to the same limit as for the exact optimal filter.

MSC: Primary 60G15, 60G35. Secondary 62M20, 93E11

Keywords: fractional Brownian motion, homogeneous linear system, optimal filtering, filtering error, asymptotic variance

1 Introduction

Several contributions have been already reported around filtering problems concerning models where the driving processes are fractional Brownian motions (fBm's for short) : see Kleptsyna *et al.* (2000) for a rather general approach and further references. The specific case of a homogeneous linear system has been investigated in Kleptsyna and Le Breton (2002) where explicit closed form equations are derived both for the optimal

---

⋆ Author's research was supported by the Grant INTAS 99-00559. She is also associated with the Institute of Information Transmission Problems, Moscow, Russia.

* Laboratoire de Statistique et Processus/Université du Maine. Av. Olivier Messiaen, 72085 Le Mans, Cedex 9, France. e-mail: marina.kleptsyna@univ-lemans.fr.

** *Address for correspondence:* Fax: 33 476 631 263. His research was supported by the Project IDOPT, CNRS-UJF-INPG-INRIA. Laboratoire de Modélisation et Calcul/Université J. Fourier. BP 53, 38041 Grenoble Cedex 9, France. e-mail: alain.le-breton@imag.fr.

*** Laboratoire de Modélisation et Calcul/Université J. Fourier. BP 53, 38041 Grenoble Cedex 9, France. e-mail: a-c.fouilhe-viot@wanadoo.fr.

Received: October 2003
Accepted: May 2004

filter and the variance of the filtering error. Moreover, therein it is shown that this filter is asymptotically stable in the sense that the variance of the filtering error converges to a finite limit as the observation time tends to infinity. Here our aim is to exhibit a simple approximate filter which has the same asymptotic behaviour as the optimal one. Let us fix this more precisely.

As in Kleptsyna and Le Breton (2002), we deal with real-valued processes $X = (X_t, t \geq 0)$ and $Y = (Y_t, t \geq 0)$, representing the signal and the observation respectively, governed by the following homogeneous linear system of stochastic differential equations interpreted as integral equations :

$$\begin{cases} dX_t &= \theta X_t dt + dV_t^H, \ t \geq 0, \ X_0 = 0, \\ dY_t &= \mu X_t dt + dW_t^H, \ t \geq 0, \ Y_0 = 0. \end{cases} \tag{1.1}$$

Here $V^H = (V_t^H, t \geq 0)$ and $W^H = (W_t^H, t \geq 0)$ are independent normalized fBm's with the same Hurst parameter H in $[\frac{1}{2}, 1)$ and the coefficients θ and $\mu \neq 0$ are fixed real constants. The system (1.1) has a uniquely defined solution process (X, Y) which is Gaussian. Supposing that only Y is observed but one wishes to know X, the classical problem of filtering the signal X at time t from the observation of Y up to time t occurs. The solution to this problem is the conditional distribution of X_t given $\{Y_s, 0 \leq s \leq t\}$, which of course is Gaussian. Then, it is completely determined by the conditional mean $\pi_t(X) = \mathbf{E}(X_t/\{Y_s, 0 \leq s \leq t\})$, which we shall call the *exact optimal filter*, and the variance $\gamma_{xx}(t) = \mathbf{E}(X_t - \pi_t(X))^2$ of the filtering error. In Kleptsyna and Le Breton (2002), a system of Volterra type integral equations for these characteristics is provided and the following stability property of the filter is also shown :

$$\lim_{t \to +\infty} \gamma_{xx}(t) = \gamma_H,$$

where the constant γ_H is given by

$$\gamma_H = \frac{\Gamma(2H+1)}{2(\theta^2 + \mu^2)^H}[1 + \frac{\sqrt{\theta^2+\mu^2} + \theta}{\sqrt{\theta^2+\mu^2} - \theta} \sin \pi H]. \tag{1.2}$$

In the classical case $H = \frac{1}{2}$ where the noises are standard Brownian motions, the system of filtering equations reduces to the well-known Kalman-Bucy system (see, *e.g.*, Davis (1977) and Liptser and Shiryaev (1978)) and the asymptotic variance of the filtering error is $\gamma_{\frac{1}{2}} = \mu^{-2}[\sqrt{\theta^2 + \mu^2} + \theta]$. In that case, substituting the constant $\gamma_{\frac{1}{2}}$ for the function $\gamma_{xx}(t)$ in the Kalman-Bucy system, one gets the simpler filtering equation

$$d\pi_t^*(X) = -\sqrt{\theta^2 + \mu^2}\pi_t^*(X)dt + \mu\gamma_{\frac{1}{2}} dY_t; \ \pi_0^*(X) = 0, \tag{1.3}$$

which generates the filter

$$\pi_t^*(X) = \mu\gamma_{\frac{1}{2}} \int_0^t e^{-\sqrt{\theta^2+\mu^2}(t-s)} dY_s. \tag{1.4}$$

It turns out that $\pi_t^*(X)$ is an *asymptotically optimal filter* in the sense that the variance $\mathbf{E}(X_t - \pi_t^*(X))^2$ of the corresponding filtering error converges to $\gamma_{\frac{1}{2}}$ as t goes to infinity. Observe that actually, in this case, the asymptotic optimality in filtering is achieved in the class of filters which can be represented as $\int_0^t \phi(t-s)dY_s$. In the present paper, we show that this still holds for $H > \frac{1}{2}$ and we identify in this class a filter for which the variance of the filtering error converges to γ_H.

The paper is organized as follows. At first in Section 2, we fix some notations and preliminaries; in particular we associate to the problem under study an equivalent deterministic control problem. Then, our main result is stated and proved in Section 3 by exploiting the solution of this auxiliary problem which belongs to a family of infinite time horizon deterministic control problems which are investigated in Section 4.

2 Preliminaries

Fractional Brownian motion. Here, for some $H \in [\frac{1}{2}, 1)$, $B^H = (B_t^H, t \geq 0)$ is a normalized fractional Brownian motion with Hurst parameter H. This means that B^H is a Gaussian process with continuous paths such that $B_0^H = 0$, $\mathbf{E}B_t^H = 0$ and

$$\mathbf{E}B_s^H B_t^H = \frac{1}{2}[s^{2H} + t^{2H} - |s-t|^{2H}], \quad s, t \geq 0. \tag{2.1}$$

Of course the fBm reduces to the standard Brownian motion when $H = \frac{1}{2}$. For $H \neq \frac{1}{2}$, the fBm is outside the world of semimartingales but a theory of stochastic integration with respect to fBm has been developed (see, *e.g.*, Decreusefond and Üstünel (1999) or Duncan *et al.* (2000)). Actually the case of deterministic integrands, which is sufficient for the purpose of the present paper, is easy to handle (see, *e.g.*, Norros *et al.* (1999)). In particular, for a stochastic integral

$$S_t = \int_0^t g(t-s)dB_s^H, \tag{2.2}$$

we can evaluate

$$\mathbf{E}S_t^2 = \begin{cases} \displaystyle\int_0^t g^2(s)ds & \text{if } H = \frac{1}{2}, \\ \displaystyle H(2H-1)\int_0^t \int_0^t g(s)g(r)\,|s-r|^{2H-2}\,dsdr & \text{if } H \in (\frac{1}{2}, 1). \end{cases}$$

In the second case, exploiting the representation

$$|s-r|^{2H-2} = \frac{1}{B(H-\frac{1}{2}, 2-2H)} \int_{s\vee r}^{+\infty} (\tau - s)^{H-\frac{3}{2}}(\tau - r)^{H-\frac{3}{2}}d\tau,$$

where $B(.,.)$ denotes the Beta function, it is easy to check that we can rewrite

$$\mathbf{E}S_t^2 = \frac{H(2H-1)}{B(H-\frac{1}{2}, 2-2H)} \int_0^{+\infty} \{\int_0^{s\wedge t} g(r)(s-r)^{H-\frac{3}{2}}dr\}^2 ds.$$

Therefore, we have also for all $H \in [\frac{1}{2}, 1)$

$$\lim_{t \to +\infty} \mathbf{E} S_t^2 = \frac{2H\Gamma(\frac{3}{2} - H)}{\Gamma(H + \frac{1}{2})\Gamma(2 - 2H)} \int_0^{+\infty} \widetilde{g}^2(s) ds, \tag{2.3}$$

where

$$\widetilde{g}(s) = \frac{d}{ds} \int_0^s g(r)(s - r)^{H - \frac{1}{2}} dr, \tag{2.4}$$

and Γ is the Gamma function. Actually, the connection (2.4) can be inverted by

$$g(s) = \frac{1}{B(H + \frac{1}{2}, \frac{3}{2} - H)} \frac{d}{ds} \int_0^s \widetilde{g}(r)(s - r)^{\frac{1}{2} - H} dr. \tag{2.5}$$

Filtering errors. As announced in Section 1, in the system (1.1), we shall concentrate on filters which take the form

$$\pi_t^\phi(X) = \int_0^t \phi(t - s) dY_s .$$

From the first equation in (1.1), we have

$$X_t = \mathrm{e}^{\theta t} \int_0^t \mathrm{e}^{-\theta s} dV_s^H ,$$

and, taking into account the second one, we get

$$\pi_t^\phi(X) = \mu \int_0^t \phi(t - s) \mathrm{e}^{\theta s} \{ \int_0^s \mathrm{e}^{-\theta u} dV_u^H \} ds + \int_0^t \phi(t - s) dW_s^H .$$

Hence, it comes that

$$\pi_t^\phi(X) = \mu \int_0^t \{ \int_u^t \phi(t - s) \mathrm{e}^{\theta s} ds \} \mathrm{e}^{-\theta u} dV_u^H + \int_0^t \phi(t - s) dW_s^H ,$$

or

$$\pi_t^\phi(X) = \mu \int_0^t \{ \int_0^{t-u} \phi(w) \mathrm{e}^{-\theta w} dw \} \mathrm{e}^{\theta(t-u)} dV_u^H + \int_0^t \phi(t - s) dW_s^H .$$

Finally, the filtering error corresponding to the filter $\pi_t^\phi(X)$ can be written as

$$X_t - \pi_t^\phi(X) = \int_0^t \mathrm{e}^{\theta(t-s)} \{ 1 - \mu \int_0^{t-s} \phi(w) \mathrm{e}^{-\theta w} dw \} dV_s^H - \int_0^t \phi(t - s) dW_s^H ,$$

or equivalently

$$X_t - \pi_t^\phi(X) = \int_0^t Z^\phi(t - s) dV_s^H - \int_0^t \phi(t - s) dW_s^H , \tag{2.6}$$

where the function Z^ϕ is defined from ϕ, $Z^\phi = Z$ say, by

$$Z(\tau) = e^{\theta\tau}\{1 - \mu \int_0^\tau \phi(w)e^{-\theta w}dw\}.$$

Notice that Z is governed by the differential equation

$$\dot{Z}(\tau) = \theta Z(\tau) - \mu\phi(\tau)\,;\; Z(0) = 1\,. \tag{2.7}$$

Asymptotic variance of filtering errors. Now, starting from (2.6), according to the identities (2.2)-(2.4) with (Z, V^H) and (ϕ, W^H) in place of (g, B^H) and due to the independence of V^H and W^H, we get that the asymptotic variance of the filtering error corresponding to the filter $\pi_t^\phi(X)$, *i.e.*,

$$\lim_{t\to+\infty} \mathbf{E}(X_t - \pi_t^\phi(X))^2 = J(\phi)\,, \tag{2.8}$$

is given by

$$J(\phi) = \frac{2H\Gamma(\frac{3}{2} - H)}{\Gamma(H + \frac{1}{2})\Gamma(2 - 2H)} \int_0^{+\infty} \{\widetilde{Z}^2(s) + \widetilde{\phi}^2(s)\}ds\,, \tag{2.9}$$

where, for Z linked to ϕ by (2.7),

$$\widetilde{Z}(s) = \frac{d}{ds}\int_0^s Z(r)(s - r)^{H-\frac{1}{2}}dr\,;\quad \widetilde{\phi}(s) = \frac{d}{ds}\int_0^s \phi(r)(s - r)^{H-\frac{1}{2}}dr\,. \tag{2.10}$$

Actually, it is readily seen from (2.7) and (2.10) that the dynamics which links $\widetilde{Z}$ to $\widetilde{\phi}$ is nothing but

$$\widetilde{Z}(t) = \theta \int_0^t \widetilde{Z}(s)ds - \mu \int_0^t \widetilde{\phi}(s)ds + t^{H-\frac{1}{2}}\,. \tag{2.11}$$

Notice that of course if $H = \frac{1}{2}$, and hence $\widetilde{\phi} \equiv \phi$ and $\widetilde{Z} \equiv Z$, equation (2.11) is nothing but equation (2.7) written in integral form and if $H > \frac{1}{2}$, then (2.11) can be rewritten as

$$\dot{\widetilde{Z}}(t) = \theta\widetilde{Z}(s) - \mu\widetilde{\phi}(s)ds + (H - \frac{1}{2})t^{H-\frac{3}{2}}\,;\quad \widetilde{Z}(0) = 0\,.$$

Due to the limiting property (2.8), our guess is that in order to define an asymptotically optimal filter $\pi_t^*(X)$, one may take $\pi_t^*(X) = \pi_t^{\phi^*}(X)$ where the function ϕ^* corresponds through (2.5) to an optimal control $\widetilde{\phi}^*$ in the control problem :

$$\min_{\widetilde{\phi}} \widetilde{J}(\widetilde{\phi}) \quad \text{subject to (2.11)}\,, \tag{2.12}$$

with the performance criterion $\widetilde{J}(\widetilde{\phi}) = J(\phi)$ defined by (2.9).

The concerned infinite time horizon deterministic control problem (2.12) belongs to the class of control problems which are solved in Section 4. Their solutions make us able to formulate and prove our main result.

3 Asymptotically optimal filtering

At first, let us discuss the case when $H = \frac{1}{2}$. Here, in the control problem studied in Section 4, we must take $x = 1$, $K \equiv 0$, $a = \theta$, $b = -\mu$ and $q = r = 1$. Hence, applying Theorem 4.1 (see also the particular case 4.1), it comes that the optimal control in (2.12) is

$$\phi^*(t) = \mu\gamma_{\frac{1}{2}} e^{-\sqrt{\theta^2+\mu^2}t},$$

where

$$\gamma_{\frac{1}{2}} = \frac{\sqrt{\theta^2+\mu^2}+\theta}{\mu^2},$$

is the value of the optimal cost. This means nothing but that, as claimed in Section 1, an asymptotically optimal filter is $\pi_t^*(X) = \pi_t^{\phi^*}(X)$ given by (1.4).

Now, we turn to the case $H \in (\frac{1}{2}, 1)$ where we can prove the following statement which provides also an asymptotically optimal filter :

Theorem 3.1 *Define the function* V^* *by*

$$V^*(t) = \frac{H-\frac{1}{2}}{B(H+\frac{1}{2}, \frac{3}{2}-H)} \int_0^{+\infty} e^{-\sqrt{\theta^2+\mu^2}t\tau} \frac{\tau^{H-\frac{1}{2}}}{\tau+1} d\tau, \quad t > 0. \tag{3.1}$$

Let the pair of functions (ϕ^*, Z^*) *be defined by*

$$\begin{cases} \phi^*(t) &= \dfrac{\theta+\sqrt{\theta^2+\mu^2}}{\mu}[Z^*(t)+V^*(t)], \\ \dot{Z}^*(t) &= \theta Z^*(t) - \mu\phi^*(t); \quad Z^*(0) = 1. \end{cases} \tag{3.2}$$

Then the filter

$$\pi_t^*(X) = \int_0^t \phi^*(t-s)dY_s,$$

is asymptotically optimal, i.e.,

$$\lim_{t\to+\infty} \boldsymbol{E}(X_t - \pi_t^*(X))^2 = \gamma_H,$$

where γ_H *is given by (1.2).*

Proof. For $H \in (\frac{1}{2}, 1)$, in the control problem studied in Section 4, we must take $x = 0$, $K(t) = (H-\frac{1}{2})t^{H-\frac{3}{2}}$, $a = \theta$, $b = -\mu$ and

$$q = r = \frac{2H\Gamma(\frac{3}{2}-H)}{\Gamma(H+\frac{1}{2})\Gamma(2-2H)}.$$

Hence, applying Theorem 4.1 (see also the particular case 4.2), we get that the following pair $(\widetilde{\phi}^*, \widetilde{Z}^*)$ is optimal in the control problem (2.12) :

$$\begin{cases} \widetilde{\phi}^*(t) &= \dfrac{\theta+\sqrt{\theta^2+\mu^2}}{\mu}[\widetilde{Z}^*(t)+\widetilde{V}^*(t)], \\ \dot{\widetilde{Z}}^*(t) &= \theta\widetilde{Z}^*(t) - \mu\widetilde{\phi}^*(t) + (H-\frac{1}{2})t^{H-\frac{3}{2}}; \quad \widetilde{Z}^*(0) = 0, \end{cases}$$

where

$$\widetilde{V}^*(t) = (H - \frac{1}{2}) \int_0^{+\infty} e^{-\sqrt{\theta^2+\mu^2}r}(t+r)^{H-\frac{3}{2}} dr. \tag{3.3}$$

Moreover, it is easy to check that the optimal cost in (2.12) is $\widetilde{J}(\widetilde{\phi}^*) = \gamma_H$ where γ_H is given by (1.2). Hence, it is clear that to define an asymptotically optimal filter by $\pi_t^*(X) = \pi_t^{\phi^*}(X)$ we can take the second component ϕ^* of the triple (V^*, ϕ^*, Z^*) which corresponds through (2.5) to the triple $(\widetilde{V}^*, \widetilde{\phi}^*, \widetilde{Z}^*)$. It is easy to check that ϕ^* is defined by (3.2) where V^* corresponds through (2.5) to $\widetilde{V}^*$ and so, finally, we have just to identify V^*. From (3.3), we compute

$$\begin{aligned}
\int_0^t (t-s)^{\frac{1}{2}-H}\widetilde{V}^*(s)ds &= (H-\frac{1}{2})\int_0^t (t-s)^{\frac{1}{2}-H}\{\int_0^{+\infty} e^{-\sqrt{\theta^2+\mu^2}r}(s+r)^{H-\frac{3}{2}}dr\}ds \\
&= (H-\frac{1}{2})\int_0^{+\infty} e^{-\sqrt{\theta^2+\mu^2}r}\{\int_0^t (t-s)^{\frac{1}{2}-H}(s+r)^{H-\frac{3}{2}}ds\}dr \\
&= (H-\frac{1}{2})\int_0^{+\infty} e^{-\sqrt{\theta^2+\mu^2}r}\{\int_0^{\frac{t}{t+r}} v^{\frac{1}{2}-H}(1-v)^{H-\frac{3}{2}}dv\}dr.
\end{aligned}$$

Observing that actually

$$\frac{d}{dt}\int_0^{\frac{t}{t+r}} v^{\frac{1}{2}-H}(1-v)^{H-\frac{3}{2}}dv = \frac{t^{\frac{1}{2}-H}r^{H-\frac{1}{2}}}{t+r},$$

it follows that

$$\begin{aligned}
\int_0^t (t-s)^{\frac{1}{2}-H}\widetilde{V}^*(s)ds &= (H-\frac{1}{2})\int_0^{+\infty} e^{-\sqrt{\theta^2+\mu^2}r}\{\int_0^t \frac{u^{\frac{1}{2}-H}r^{H-\frac{1}{2}}}{u+r}du\}dr \\
&= (H-\frac{1}{2})\int_0^t\{\int_0^{+\infty} e^{-\sqrt{\theta^2+\mu^2}r}\frac{u^{\frac{1}{2}-H}r^{H-\frac{1}{2}}}{u+r}dr\}du \\
&= (H-\frac{1}{2})\int_0^t\{\int_0^{+\infty} e^{-\sqrt{\theta^2+\mu^2}u\tau}\frac{\tau^{H-\frac{1}{2}}}{\tau+1}d\tau\}du.
\end{aligned}$$

From (2.5), we see that this means exactly that V^* is given by (3.1). □

Remark 3.1 (*a*) Observe that from (3.1) we have also

$$\dot{V}^*(t) = -\sqrt{\theta^2+\mu^2}\frac{H-\frac{1}{2}}{B(H+\frac{1}{2},\frac{3}{2}-H)}\int_0^{+\infty} e^{-\sqrt{\theta^2+\mu^2}t\tau}\frac{\tau^{H+\frac{1}{2}}}{\tau+1}d\tau, \quad t>0.$$

Then, splitting the integral into two terms corresponding to the decomposition of $\tau^{H+\frac{1}{2}}$ as the difference $[\tau^{H+\frac{1}{2}} + \tau^{H-\frac{1}{2}}] - \tau^{H-\frac{1}{2}}$, one may easily check that V^* is actually the solution of the differential equation

$$\dot{V}^*(t) = \sqrt{\theta^2+\mu^2}V^*(t) - \beta_H(H-\frac{1}{2})t^{-\frac{1}{2}-H}; \quad \lim_{t\to+\infty} V^*(t) = 0,$$

where

$$\beta_H = \frac{(\theta^2+\mu^2)^{\frac{1-2H}{4}}}{\Gamma(\frac{3}{2}-H)}.$$

Since $\int_0^{+\infty} V^*(t)dt = 1$, it means also that that V^* is the solution of the integral equation

$$V^*(t) = \int_0^t \sqrt{\theta^2 + \mu^2 V^*(s)ds} + \beta_H t^{\frac{1}{2}-H} - 1 . \tag{3.4}$$

(*b*) Let us emphasize that, similarly to the case $H = \frac{1}{2}$ where the filter $\pi_t^*(X)$ can be generated by the approximate Kalman-Bucy algorithm (1.3), a recursive scheme can be also provided for the asymptotically optimal filter in the case $H \in (\frac{1}{2}, 1)$. At first, we observe that due to the first equation in (3.2) we can write

$$\pi_t^*(X) = \mu\gamma_{\frac{1}{2}}[\mathcal{Z}_t^* + \mathcal{V}_t^*] , \tag{3.5}$$

where

$$\mathcal{Z}_t^* = \int_0^t Z^*(t-s)dY_s ; \quad \mathcal{V}_t^* = \int_0^t V^*(t-s)dY_s .$$

Since the function Z^* is differentiable, we have

$$\mathcal{Z}_t^* = Z^*(0)Y_t + \int_0^t \{\int_0^s \dot{Z}^*(s-r)dY_r\}ds .$$

Hence, due to the second equation in (3.2), the process $\mathcal{Z}^*$ is generated from Y by the equation

$$\mathcal{Z}_t^* = \theta \int_0^t \mathcal{Z}_s^* ds - \mu \int_0^t \pi_s^*(X)ds + Y_t , \tag{3.6}$$

Now, using equation (3.4), we can write

$$\mathcal{V}_t^* = \int_0^t \psi(t-s)dY_s + \beta_H \int_0^t (t-s)^{\frac{1}{2}-H} dY_s ,$$

where the function ψ satisfies

$$\dot{\psi}(t) = \sqrt{\theta^2 + \mu^2 V^*(t)} ; \quad \psi(0) = -1 .$$

Consequently, we get that

$$\begin{aligned} \int_0^t \psi(t-s)dY_s &= \psi(0)Y_t + \int_0^t \{\int_0^s \dot{\psi}(s-r)dY_r\}ds \\ &= \sqrt{\theta^2 + \mu^2} \int_0^t \mathcal{V}_s^* ds - Y_t . \end{aligned}$$

Finally, the following equation holds for $\mathcal{V}_t^*$:

$$\mathcal{V}_t^* = \sqrt{\theta^2 + \mu^2} \int_0^t \mathcal{V}_s^* ds + \int_0^t [\beta_H (t-s)^{\frac{1}{2}-H} - 1] dY_s . \tag{3.7}$$

The system (3.5)-(3.7) provides a closed-form recursion which generates the filter $\pi_t^*(X)$ from the observation process Y. It is readily seen that when $H = \frac{1}{2}$, and hence $\mathcal{V}^* \equiv 0$ and $\pi_t^*(X) = \mu\gamma_{\frac{1}{2}}\mathcal{Z}_t^*$, this system reduces to the single equation

$$\pi_t^*(X) = -\sqrt{\theta^2+\mu^2}\int_0^t \pi_s^*(X)ds + \mu\gamma_{\frac{1}{2}}Y_t\,,$$

which is nothing but equation (1.3). (*c*) Suppose that $H > \frac{1}{2}$ but one does as if the noises were standard Brownian motions and hence uses the filter generated by the approximate Kalman-Bucy algorithm (1.3), *i.e.*, the filter

$$\widetilde{\pi}_t(X) = \mu\gamma_{\frac{1}{2}}\int_0^t e^{-\sqrt{\theta^2+\mu^2}(t-s)}dY_s\,.$$

Then it can be checked that the corresponding asymptotic variance of the filtering error $\lim_{t\to+\infty}\mathbf{E}(X_t - \widetilde{\pi}_t(X))^2$ is the constant

$$\widetilde{\gamma}_H = \frac{\Gamma(2H+1)}{(\theta^2+\mu^2)^{H-\frac{1}{2}}}\gamma_{\frac{1}{2}}\,.$$

Moreover the consequent loss of performance with respect to the asymptotically optimal filter can be evaluated by

$$\widetilde{\gamma}_H - \gamma_H = \frac{\Gamma(2H+1)}{2(\theta^2+\mu^2)^H}\mu^2\gamma_{\frac{1}{2}}^2(1-\sin\pi H)\,.$$

Let us observe that, for fixed parameters θ and μ, the asymptotic relative efficiency

$$\frac{\gamma_H}{\widetilde{\gamma}_H} = \frac{1+\mu^2\gamma_{\frac{1}{2}}^2\sin\pi H}{1+\mu^2\gamma_{\frac{1}{2}}^2}\,,$$

of $\widetilde{\pi}_t(X)$ decreases as H increases in $(\frac{1}{2}, 1)$.

4 About optimal control problems

Given a function $K = (K(t), t \geq 0)$ and constants a and b, we consider the state dynamics

$$\dot{\mathcal{X}}_t = a\mathcal{X}_t + b\mathcal{U}_t + K(t)\,,\ t\geq 0;\quad \mathcal{X}_0 = x\,, \tag{4.1}$$

where the control $\mathcal{U} = (\mathcal{U}_t, t \geq 0)$ can be chosen in order to drive the state $\mathcal{X} = (\mathcal{X}_t, t \geq 0)$. Let $\mathcal{A}$ be the class of measurable functions $\mathcal{U}$, called admissible controls, such that the corresponding differential equation (4.1) has a unique solution $\mathcal{X}$. Given constants $q > 0$ and $r > 0$, we define the performance criterion $\mathcal{J}$ by

$$\mathcal{J}(\mathcal{U}) = \int_0^{+\infty}[q\mathcal{X}_t^2 + r\mathcal{U}_t^2]dt\,. \tag{4.2}$$

The following statement gives the solution of the infinite time horizon deterministic control problem corresponding to (4.1)-(4.2).

Theorem 4.1 *Define the constants*

$$\rho = \frac{r}{b^2}[a + \delta]\,; \quad \delta = \sqrt{a^2 + \frac{b^2}{r}q}\,. \tag{4.3}$$

Assume that $\lim_{t\to+\infty} K(t) = 0$ *and also, setting*

$$\mathcal{V}_K(t) = \int_0^{+\infty} e^{-\delta r} K(t + r)dr\,,\ t \geq 0\,, \tag{4.4}$$

that the function $\mathcal{V}_K$ *is well-defined. Let the pair* $(\mathcal{U}^*, \mathcal{X}^*)$ *be governed by*

$$\begin{cases} \mathcal{U}_t^* &= -\frac{b}{r}\rho[\mathcal{X}_t^* + \mathcal{V}_K(t)]\,, \\ \dot{\mathcal{X}}_t^* &= a\mathcal{X}_t^* + b\mathcal{U}_t^* + K(t)\,; \quad \mathcal{X}_0^* = x\,, \end{cases} \tag{4.5}$$

Then, for $\mathcal{J}$ *defined by (4.2), the pair* $(\mathcal{U}^*, \mathcal{X}^*)$ *is optimal in the control problem*

$$\min_{\mathcal{U}\in\mathcal{A}} \mathcal{J}(\mathcal{U}) \quad \text{subject to (4.1)}\,.$$

Moreover, the value of the optimal cost is

$$\mathcal{J}(\mathcal{U}^*) = \rho[x + \mathcal{V}_K(0)]^2 + q\int_0^{+\infty} \mathcal{V}_K^2(s)ds\,. \tag{4.6}$$

Proof. Suppose that there exists a pair $(\mathcal{X}^*, p^*)$ which satisfies the Hamiltonian system

$$\begin{aligned} \dot{\mathcal{X}}_t^* &= a\mathcal{X}_t^* - \frac{b^2}{r}p_t^* + K(t)\,; \quad \mathcal{X}_0^* = x\,, \\ \dot{p}_t^* &= -q\mathcal{X}_t^* - ap_t^*\,; \quad \lim_{t\to+\infty} p_t^* = 0\,, \end{aligned} \tag{4.7}$$

Hence of course $\mathcal{X}^*$ is nothing but the state dynamics corresponding through (4.1) to the control $\mathcal{U}^*$ defined by $\mathcal{U}_t^* = -(b/r)p_t^*$. Let us show that for an arbitrary control $\mathcal{U} \in \mathcal{A}$ the inequality $\mathcal{J}(\mathcal{U}) \geq \mathcal{J}(\mathcal{U}^*)$ holds. Of course it is true when $\mathcal{J}(\mathcal{U}) = +\infty$ and so we concentrate on the case when $\mathcal{J}(\mathcal{U}) < +\infty$ which in particular means that $\lim_{t\to+\infty} \mathcal{X}_t = 0$ for the corresponding state dynamics $\mathcal{X}$. Defining for $T > 0$

$$\mathcal{J}_T(\mathcal{U}) = \int_0^T [q\mathcal{X}_t^2 + r\mathcal{U}_t^2]dt\,, \tag{4.8}$$

we evaluate

$$\mathcal{J}_T(\mathcal{U}) = \mathcal{J}_T(\mathcal{U}^*) + \int_0^T \{q[\mathcal{X}_t^2 - (\mathcal{X}_t^*)^2] + r[\mathcal{U}_t^2 - (\mathcal{U}_t^*)^2]\}dt\,.$$

Using the equality $y^2 - (y^*)^2 = (y - y^*)^2 + 2y^*(y - y^*)$ and exploiting the property $\mathcal{U}_t^* = -(b/r)p_t^*$, it is readily seen that

$$\mathcal{J}_T(\mathcal{U}) = \mathcal{J}_T(\mathcal{U}^*) + \Delta_1(T) + 2\Delta_2(T), \tag{4.9}$$

where

$$\begin{aligned} \Delta_1(T) &= \int_0^T \{q[\mathcal{X}_t - \mathcal{X}_t^*]^2 + r[\mathcal{U}_t - \mathcal{U}_t^*]^2\}dt, \\ \Delta_2(T) &= \int_0^T \{q\mathcal{X}_t^*[\mathcal{X}_t - \mathcal{X}_t^*] - bp_t^*[\mathcal{U}_t - \mathcal{U}_t^*]\}dt. \end{aligned}$$

But, rewriting the quantity in the last integral as

$$(\mathcal{X}_t - \mathcal{X}_t^*)[q\mathcal{X}_t^* + ap_t^*] - p_t^*[a(\mathcal{X}_t - \mathcal{X}_t^*) + b(\mathcal{U}_t - \mathcal{U}_t^*)],$$

and taking into account equations (4.1) and (4.7), we see that this integral can be written as

$$-\int_0^T (\mathcal{X}_t - \mathcal{X}_t^*)dp_t^* - \int_0^T p_t^* d(\mathcal{X}_t - \mathcal{X}_t^*).$$

Therefore, integrating by parts, since $\mathcal{X}_0 - \mathcal{X}_0^* = 0$, it comes that

$$\Delta_2(T) = -p_T^*(\mathcal{X}_T - \mathcal{X}_T^*).$$

Consequently, since $\Delta_1(T) \geq 0$, from (4.9) we get that

$$\mathcal{J}_T(\mathcal{U}) \geq \mathcal{J}_T(\mathcal{U}^*) - 2p_T^*(\mathcal{X}_T - \mathcal{X}_T^*).$$

Hence, if $\lim_{T\to+\infty} \mathcal{X}_T^* = 0$, letting T tend to infinity in this inequality, due to the limiting conditions for p^* and $\mathcal{X}$, we obtain $\mathcal{J}(\mathcal{U}) \geq \mathcal{J}(\mathcal{U}^*)$.
Now, to show that the pair $(\mathcal{U}^*, \mathcal{X}^*)$ defined by (4.5) is optimal, it is sufficient to check that the pair $(\mathcal{X}^*, p^*)$, where $p_t^* = \rho[\mathcal{X}_t^* + \mathcal{V}_K(t)]$, satisfies the Hamiltonian system (4.7) and that also the limiting condition $\lim_{t\to+\infty} \mathcal{X}_t^* = 0$ holds. At first, it is easy to check that $(\mathcal{X}^*, p^*)$ satisfies the differential equations in (4.7). One can observe that the expression (4.4) for $\mathcal{V}_K$ can be rewritten as

$$\mathcal{V}_K(t) = \int_t^{+\infty} e^{\delta(t-s)} K(s)ds, \ t \geq 0,$$

and since $\lim_{t\to+\infty} K(t) = 0$, actually $\mathcal{V}_K$ is nothing but the solution of the equation

$$\dot{\mathcal{V}}_K(t) = \delta\mathcal{V}_K(t) - K(t); \quad \lim_{t\to+\infty} \mathcal{V}_K(t) = 0. \tag{4.10}$$

Now, since from the first equation in (4.7) we have

$$\dot{\mathcal{X}}_t^* = -\delta\mathcal{X}_t^* - \frac{b^2}{r}\rho\mathcal{V}_K(t) + K(t); \quad \mathcal{X}_0^* = x, \tag{4.11}$$

due to $\lim_{t\to+\infty} K(t) = \lim_{t\to+\infty} \mathcal{V}_K(t) = 0$, it is clear that $\lim_{t\to+\infty} \mathcal{X}_t^* = 0$. Hence, we have also $\lim_{t\to+\infty} p_t^* = 0$.

Finally, we evaluate the optimal cost $\mathcal{J}(\mathcal{U}^*)$. At first, in order to compute the variation $p_T^*\mathcal{X}_T^* - p_0^*\mathcal{X}_0^*$, we express $p_t^*\dot{\mathcal{X}}_t^* + \dot{p}_t^*\mathcal{X}_t^*$ from (4.7). Then, for $\mathcal{J}_T$ defined by (4.8), it follows easily that

$$\mathcal{J}_T(\mathcal{U}^*) = p_0^*x - p_T^*\mathcal{X}_T^* + \int_0^T p_t^*K(t)dt.$$

Hence, since $p_t^* = \rho[\mathcal{X}_t^* + \mathcal{V}_K(t)]$, taking the limit for T tending to infinity, we get

$$\mathcal{J}(\mathcal{U}^*) = \rho[x + \mathcal{V}_K(0)]x + \rho\int_0^{+\infty} [\mathcal{X}_t^* + \mathcal{V}_K(t)]K(t)dt.$$

Proceeding similarly through the evaluation of the variation $\mathcal{V}_K(T)\mathcal{X}_T^* - \mathcal{V}_K(0)\mathcal{X}_0^*$ from (4.10)-(4.11), we obtain that

$$\int_0^{+\infty} \mathcal{X}_t^*K(t)dt = -\frac{b^2}{r}\rho\int_0^{+\infty} \mathcal{V}_K^2(t)dt + \int_0^{+\infty} K(t)\mathcal{V}_K(t)dt + \mathcal{V}_K(0)x.$$

Then, since from equation (4.10) we have $K(t) = \delta\mathcal{V}_K(t) - \dot{\mathcal{V}}_K(t)$, it follows that

$$\mathcal{J}(\mathcal{U}^*) = \rho[x^2 + 2\mathcal{V}_K(0)x] + \rho(\delta - a)\int_0^{+\infty} \mathcal{V}_K^2(t)dt - 2\rho\int_0^{+\infty} \mathcal{V}_K(t)\dot{\mathcal{V}}_K(t)dt.$$

But $\rho(\delta - a) = q$ and clearly the last integral equals $-\frac{1}{2}\mathcal{V}_K^2(0)$ and so the equality (4.6) holds. □

Remark 4.1 Actually, from (4.10), we observe that

$$\mathcal{V}_K(t)\dot{\mathcal{V}}_K(t) = \delta\mathcal{V}_K^2(t) - K(t)\mathcal{V}_K(t),$$

and hence

$$\int_0^{+\infty} \mathcal{V}_K^2(t)dt = \frac{1}{\delta}\{\int_0^{+\infty} K(t)\mathcal{V}_K(t)dt - \frac{1}{2}\mathcal{V}_K^2(0)\}.$$

This allows to rewrite the value (4.6) of the optimal cost as

$$\mathcal{J}(\mathcal{U}^*) = \rho x[x + 2\mathcal{V}_K(0)] + \frac{\rho}{\delta}\{\frac{\delta + a}{2}\mathcal{V}_K^2(0) + (\delta - a)\int_0^{+\infty} K(t)\mathcal{V}_K(t)dt\}. \quad (4.12)$$

PARTICULAR CASE 4.1 If we take $K \equiv 0$, and hence also $\mathcal{V}_K \equiv 0$, then the optimal pair $(\mathcal{U}^*, \mathcal{X}^*)$ is governed by

$$\begin{cases} \mathcal{U}_t^* &= -\frac{b}{r}\rho\mathcal{X}_t^*, \\ \dot{\mathcal{X}}_t^* &= a\mathcal{X}_t^* + b\mathcal{U}_t^*; \quad \mathcal{X}_0^* = x. \end{cases} \quad (4.13)$$

Since $a - (b^2/r)\rho = -\delta$, this means that $\mathcal{X}_t^* = e^{-\delta t}x$ and $\mathcal{U}_t^* = -(b/r)\rho e^{-\delta t}x$. Substituting these expressions for $\mathcal{X}_t^*$ and $\mathcal{U}_t^*$ in the integral $\int_0^{+\infty}[q(\mathcal{X}_t^*)^2 + r(\mathcal{U}_t^*)^2]dt$, a direct computation gives the value $\mathcal{J}(\mathcal{U}^*) = \rho x^2$ of the optimal cost, which of course is nothing but what (4.6) says in the present case.

PARTICULAR CASE 4.2 If, for some $H \in (\frac{1}{2}, 1)$, we take $K(t) = (H - \frac{1}{2})t^{H-\frac{3}{2}}$, then the optimal pair $(\mathcal{U}^*, \mathcal{X}^*)$ is governed by (4.5) with

$$\mathcal{V}_K(t) = (H - \frac{1}{2}) \int_0^{+\infty} e^{-\delta r}(t + r)^{H-\frac{3}{2}} dr. \tag{4.14}$$

Moreover, the value of the optimal cost can be computed explicitly. Actually, here from (4.14), straightforward computations give that

$$\mathcal{V}_K(0) = \delta^{\frac{1}{2}-H}\Gamma(H + \frac{1}{2}); \quad \int_0^{+\infty} K(t)\mathcal{V}_K(t)dt = \delta^{1-2H}\frac{\Gamma(2H)\Gamma(H + \frac{1}{2})\Gamma(2 - 2H)}{2\Gamma(\frac{3}{2} - H)}.$$

Inserting this into the expression (4.12), one may finally get

$$\mathcal{J}(\mathcal{U}^*) = \rho x[x + \frac{2}{\delta^{H-\frac{1}{2}}}\Gamma(H + \frac{1}{2})] + \frac{q}{\delta^{2H}}\frac{\Gamma(2H)\Gamma(H + \frac{1}{2})\Gamma(2 - 2H)}{2\Gamma(\frac{3}{2} - H)}[1 + \frac{\delta + a}{\delta - a}\sin \pi H].$$

5 Concluding comments

Linear Quadratic Gaussian (LQG) problems concerning dynamical systems governed by Brownian motions have well-known solutions which are now quite classical. When the driving processes are fBm's, the theory is not yet completed, specially from the asymptotical point of view. In this paper, concentrating on filtering, we have illustrated the actual solvability of the problems. Actually, the infinite time horizon stochastic control problems are also tractable and in forthcoming papers we shall report the results about the regulator problem both in the case of complete and incomplete observation, the last one mixing filtering and control.

Acknowledgements

We are grateful to a reviewer for his comments, which helped to improve the presentation of this paper.

References

[1] Davis, M. H. A. (1977). *Linear Estimation and Stochastic Control*, Chapman and Hall, New York.

[2] Decreusefond, L. and Üstünel, A. S. (1999). Stochastic analysis of the fractional Brownian motion, *Potential Analysis*, 10, 177-214.

[3] Duncan, T. E., Hu, Y. and Pasik-Duncan, B. (2000). Stochastic calculus for fractional Brownian motion I. Theory, *SIAM J. Control Optimization*, 38 (2), 582-612.

[4] Kleptsyna, M. L. and Le Breton, A. (2002). Extension of the Kalman-Bucy filter to elementary linear systems with fractional Brownian noises, *Statistical Inference for Stochastic Processes*, 5 (3), 249-271.

[5] Kleptsyna, M. L., Le Breton, A. and Roubaud, M.-C. (2000). General approach to filtering with fractional Brownian noises-Application to linear systems, *Stochastics and Stochastics Reports*, 71, 119-140.

[6] Liptser, R. S. and Shiryaev, A. N. (1978). *Statistics of Random Processes*, Springer-Verlag, New York.

[7] Norros, I., Valkeila, E. and Virtamo, J. (1999). An elementary approach to a Girsanov formula and other analytical results on fractional Brownian motions, *Bernoulli*, 5 (4), 571-587.

SORT 28 (2) July-December 2004, 191-200

Estimation of the noncentrality matrix of a noncentral Wishart distribution with unit scale matrix. A matrix generalization of leung's domination result

Heinz Neudecker*

Cesaro

Abstract

The main aim is to estimate the noncentrality matrix of a noncentral Wishart distribution. The method used is Leung's but generalized to a *matrix* loss function. Parallelly Leung's scalar noncentral Wishart identity is generalized to become a matrix identity. The concept of Löwner partial ordering of symmetric matrices is used.

MSC: 62H12, 15A24, 15A45

Keywords: Noncentral Wishart matrix identity, noncentrality matrix, decision-theoretic estimation, matrix loss function, Löwner matrix ordering, Haffians

1 Introduction

We consider $S \sim W_m(n, I_m, M'M)$. Following Leung (1994) we recall that the habitual unbiased estimator of $M'M$ is $T := S - nI_m$. Under certain conditions $T_\alpha := T + \alpha(\operatorname{tr} S)^{-1} I_m$ dominates T for a suitable choice of α, as was shown by Leung, who used the loss function

$$\lambda\left[(M'M)^{-1}, R\right] := \operatorname{tr}\left\{(M'M)^{-1} R - I_m\right\}^2 .$$

He extended work by Perlman & Rasmussen (1975), Saxena & Alam (1982), Chow (1987) and Leung & Muirhead (1987).

* *Address for correspondence:* Oosterstraat, 13. 1741 GH Schagen. The Netherlands. Postal address: NL 1741 GH 13. E-mail: heinz@fee.uva.nl

Received: May 2001
Accepted: October 2003

In this article we propose to use a *matrix* loss function, *viz* $L[(M'M)^{-1}, R] := \left\{(M'M)^{-1}R - I_m\right\}'\left\{(M'M)^{-1}R - I_m\right\}$ and apply the concept of Löwner partial ordering of symmetric matrices. We shall show that Leung's result still holds approximately, the error term being of order $o(n^{-1})$. For accomplishing this we need a matrix version of Leung's Identity for the noncentral Wishart distribution. This will be presented first.

A matrix version of an ancillary lemma by Leung, *viz* his Lemma 3.1 will next be established. The generalized domination result will then follow straightforwardly.

We shall employ an approximation of $E(\mathrm{tr}\,S)^{-1}S$, where E is the expectation operator. A lemma on the matrix Haffian $\nabla\varphi F$, where φ and F are scalar and matrix functions of S, will be proved in Appendix 1. In Appendix 2 we shall prove a lemma on the scalar Haffian $\mathrm{tr}\,\nabla F_2 A F_1$, when F_1 and F_2 are matrix functions of S and A is a constant matrix.

2 A matrix version of Leung's identity for the noncentral Wishart distribution

We quote Leung's Theorem 2.1, where without loss of generality we take $h = 1$, h being a scalar function of S in Leung's work:

$$E\,\mathrm{tr}\,\Sigma^{-1}F = 2E\,\mathrm{tr}\,\nabla F + (n-m-1)E\,\mathrm{tr}\,S^{-1}F + E_1\,\mathrm{tr}\,\Sigma^{-1}M'MS^{-1}F, \tag{1}$$

where $S \sim W_m\left(n, \Sigma, \Sigma^{-1}M'M\right)$, E denotes the expectation with respect to this distribution, E_1 denotes the expectation with respect to the distribution W_m $(n+m+1, \Sigma, \Sigma^{-1}M'M)$, $F = F(S)$ and $n > m+1$. The matrices S, Σ, F and ∇ are square of dimension m, whereas M has dimension $n \times m$. It is assumed that M has full column rank. Further ∇F is the matrix Haffian as denoted by Neudecker (2000*b*). Inspired by Haff (1981), who did it for the central Wishart distribution, we shall establish a matrix version of (1).

Theorem 1

$$\begin{aligned} EF_1\Sigma^{-1}F_2 &= 2EF_1\nabla F_2 + 2\left(EF_2'\nabla F_1'\right)' + \\ &+ (n-m-1)EF_1S^{-1}F_2 + E_1F_1\Sigma^{-1}M'MS^{-1}F_2, \end{aligned} \tag{2}$$

for F_1 and F_2 satisfying the conditions of Lemma 5.

Proof. Take $F = F_2e_je_i'F_1$, with unit vectors e_i and e_j. We then use the identity:

$$\mathrm{tr}\nabla F_2AF_1 = \mathrm{tr}\,(\nabla F_2)\,AF_1 + \mathrm{tr}\,(\nabla F_1')\,A'F_2',$$

with constant A. For a proof see Lemma 5.

Taking $A = e_j e_i'$ we get

$$E\,\mathrm{tr}\Sigma^{-1}F_2 e_j e_i' F_1 = 2E\,\mathrm{tr}\,(\nabla F_2)\, e_j e_i' F_1 + 2E\,\mathrm{tr}\left(\nabla F_1'\right) e_i e_j' F_2' + \\ + (n-m-1)\mathrm{tr}\, E S^{-1} F_2 e_j e_i' F_1 + E_1\,\mathrm{tr}\,\Sigma^{-1} M'MS^{-1} F_2 e_j e_i' F_1$$

or equivalently

$$\left(EF_1\Sigma^{-1}F_2\right)_{ij} = 2\,(EF_1\nabla F_2)_{ij} + 2\left(EF_2'\nabla F_1'\right)_{ji} + \\ + (n-m-1)\left(EF_1S^{-1}F_2\right)_{ij} + \left(E_1F_1\Sigma^{-1}M'MS^{-1}F_2\right)_{ij}.$$

□

Note: It was assumed that (1) holds for all $F = F_2 e_j e_i' F_1$, which puts stronger conditions on the input matrix than was necessary for (1). By choosing $F_1 = I_m$ and taking traces we derive (1) from (2).

For discussion of the central Wishart case we refer to Haff (1981).

3 A matrix version of Leung's lemma 3.1

Lemma 2

$$E\,(tr S)^{-1}\,(M'M)^{-1}\,S\,(M'M)^{-1} < nE\,(trS)^{-1}\,(M'M)^{-2} - \\ -2(n-4)E\,(tr S)^{-2}\,(M'M)^{-2} + E_1\,(tr S)^{-1}\,(M'M)^{-1} - \\ -2E_1\,(tr S)^{-2}\,(M'M)^{-1},$$

where $S \sim W_m(n, I_m, M'M)$ and $M'M$ is assumed to be nonsingular. The inequality $A < B$, for symmetric A and B, stands for the Löwner ordering meaning that $B - A$ is positive definite.

Proof. Take $F_1 = (\mathrm{tr}\,S)^{-1}\,(M'M)^{-1}$ and $F_2 = S\,(M'M)^{-1}$. By Theorem 1 (with $\Sigma = I_m$):

$$E\,(\mathrm{tr}\,S)^{-1}\,(M'M)^{-1}\,S\,(M'M)^{-1} = 2E\,(\mathrm{tr}\,S)^{-1}\,(M'M)^{-1}\,\nabla S\,(M'M)^{-1} + \\ +2\left\{E\,(M'M)^{-1}\,S\nabla\,(\mathrm{tr}\,S)^{-1}\,(M'M)^{-1}\right\}' + \\ +(n-m-1)E\,(\mathrm{tr}\,S)^{-1}\,(M'M)^{-2} + E_1\,(\mathrm{tr}\,S)^{-1}\,(M'M)^{-1} = \\ = (m+1)E\,(\mathrm{tr}\,S)^{-1}\,(M'M)^{-2} - 2E\,(\mathrm{tr}\,S)^{-2}\,(M'M)^{-1}\,S\,(M'M)^{-1} + \\ +(n-m-1)E\,(\mathrm{tr}\,S)^{-1}\,(M'M)^{-2} + E_1\,(\mathrm{tr}\,S)^{-1}\,(M'M)^{-1} =$$

$$= nE\,(\mathrm{tr}\,S)^{-1}\,(M'M)^{-2} - 2E\,(\mathrm{tr}\,S)^{-2}\,(M'M)^{-1}\,S\,(M'M)^{-1} +$$

$$+E_1\,(\mathrm{tr}\,S)^{-1}\,(M'M)^{-1} \qquad (i)$$

Further

$$E\,(\mathrm{tr}\,S)^{-2}\,(M'M)^{-1}\,S\,(M'M)^{-1} = 2E\,(\mathrm{tr}\,S)^{-2}\,(M'M)^{-1}\,\nabla S\,(M'M)^{-1} +$$

$$+2\left\{E\,(M'M)^{-1}\,S\,\nabla\,(\mathrm{tr}\,S)^{-2}\,(M'M)^{-1}\right\}' + (n-m-1)E\,(\mathrm{tr}\,S)^{-2}\,(M'M)^{-2} +$$

$$+E_1\,(\mathrm{tr}\,S)^{-2}\,(M'M)^{-1},$$

where we applied Theorem 1 (with $\Sigma = I_m$) using $F_1 = (\mathrm{tr}\,S)^{-2}\,(M'M)^{-1}$ and $F_2 = S\,(M'M)^{-1}$. Proceeding as before we get

$$E\,(\mathrm{tr}\,S)^{-2}\,(M'M)^{-1}\,S\,(M'M)^{-1} = (m+1)E\,(\mathrm{tr}\,S)^{-2}\,(M'M)^{-2} -$$

$$-4E\,(\mathrm{tr}\,S)^{-3}\,(M'M)^{-1}\,S\,(M'M)^{-1} + (n-m-1)E\,(\mathrm{tr}\,S)^{-2}\,(M'M)^{-2} +$$

$$+E_1\,(\mathrm{tr}\,S)^{-2}\,(M'M)^{-1} = nE\,(\mathrm{tr}\,S)^{-2}\,(M'M)^{-2} -$$

$$-4E\,(\mathrm{tr}\,S)^{-3}\,(M'M)^{-1}\,S\,(M'M)^{-1} + E_1\,(\mathrm{tr}\,S)^{-2}\,(M'M)^{-1}.$$

We use the Löwner ordering $S < (\mathrm{tr}\,S)I_m$, which yields $(M'M)^{-1}\,S\,(M'M)^{-1} < (\mathrm{tr}\,S)\,(M'M)^{-2}$. Hence we get

$$(n-4)E\,(\mathrm{tr}\,S)^{-2}\,(M'M)^{-2} + E_1\,(\mathrm{tr}\,S)^{-2}\,(M'M)^{-1} < E\,(\mathrm{tr}\,S)^{-2}\,(M'M)^{-1}\,S\,(M'M)^{-1}.$$

Insertion in (i) finally yields

$$E\,(\mathrm{tr}\,S)^{-1}\,(M'M)^{-1}\,S\,(M'M)^{-1} < nE\,(\mathrm{tr}\,S)^{-1}\,(M'M)^{-2} - 2(n-4)E\,(\mathrm{tr}\,S)^{-2}\,(M'M)^{-2}$$

$$-2E_1\,(\mathrm{tr}\,S)^{-2}\,(M'M)^{-1} + E_1\,(\mathrm{tr}\,S)^{-1}\,(M'M)^{-1}.$$

□

Notes:

1. Nonsingularity of $M'M$ is not trivial. A case of singularity is $M' = \mu\iota'$, where the n means are proportional.
2. Leung assumes $n > 4$. There is no need for it.
3. Taking traces in Lemma 2 yields Leung's Lemma 3.1.

4 A matrix version of Leung's domination result

We shall now prove the main result of this paper.

Theorem 3

$$EL[(M'M)^{-1}, T] > EL[(M'M)^{-1}, T_\alpha]$$

for $0 < \alpha \leqslant 4(n-4)$, *where*

$$L[(M'M)^{-1}, R] := \left\{(M'M)^{-1} R - I_m\right\}' \left\{(M'M)^{-1} R - I_m\right\},$$

$$T := S - nI_m \quad and \quad T_\alpha := T + \alpha (tr S)^{-1} I_m.$$

Proof.

$$L[(M'M)^{-1}, T] - L[(M'M)^{-1}, T_\alpha] = \left\{(M'M)^{-1} T - I_m\right\}' \left\{(M'M)^{-1} T - I_m\right\} - \\ - \left\{(M'M)^{-1} T_\alpha - I_m\right\}' \left\{(M'M)^{-1} T_\alpha - I_m\right\} = 2n\alpha (\operatorname{tr} S)^{-1} (M'M)^{-2} - \\ -\alpha^2 (\operatorname{tr} S)^{-2} (M'M)^{-2} - \alpha (\operatorname{tr} S)^{-1} S (M'M)^{-2} - \alpha (\operatorname{tr} S)^{-1} (M'M)^{-2} S + \\ +2\alpha (\operatorname{tr} S)^{-1} (M'M)^{-1}.$$

Its expected value is

$$2n\alpha E (\operatorname{tr} S)^{-1} (M'M)^{-2} - \alpha^2 E (\operatorname{tr} S)^{-2} (M'M)^{-2} - \alpha E (\operatorname{tr} S)^{-1} S (M'M)^{-2} - \\ -\alpha E (\operatorname{tr} S)^{-1} (M'M)^{-2} S + 2\alpha E (\operatorname{tr} S)^{-1} (M'M)^{-1} > \\ > 2\alpha E (\operatorname{tr} S)^{-1} (M'M)^{-1} S (M'M)^{-1} + \\ +4\alpha(n-4) E (\operatorname{tr} S)^{-2} (M'M)^{-2} - 2\alpha E_1 (\operatorname{tr} S)^{-1} (M'M)^{-1} + 4\alpha E_1 (\operatorname{tr} S)^{-2} (M'M)^{-1} - \\ -\alpha^2 E (\operatorname{tr} S)^{-2} (M'M)^{-2} - \alpha E (\operatorname{tr} S)^{-1} S (M'M)^{-2} - \alpha E (\operatorname{tr} S)^{-1} (M'M)^{-2} S + \\ +2\alpha E (\operatorname{tr} S)^{-1} (M'M)^{-1} \\ = 2\alpha E (\operatorname{tr} S)^{-1} (M'M)^{-1} S (M'M)^{-1} + \alpha [4(n-4) - \alpha] E (\operatorname{tr} S)^{-2} (M'M)^{-2} + \\ +2\alpha \left\{E (\operatorname{tr} S)^{-1} (M'M)^{-1} - E_1 (\operatorname{tr} S)^{-1} (M'M)^{-1}\right\} + \\ +4\alpha E_1 (\operatorname{tr} S)^{-2} (M'M)^{-1} - \alpha E (\operatorname{tr} S)^{-1} S (M'M)^{-2} - \alpha E (\operatorname{tr} S)^{-1} (M'M)^{-2} S,$$

by Lemma 2.

We approximate $E (\operatorname{tr} S)^{-1} S$ by

$$\mu (n I_m + M'M) - 2\mu^2 (n I_m + 2M'M) + \\ +2\mu^3 (mn + 2 \operatorname{tr} M'M) (n I_m + M'M),$$

with $\mu^{-1} := \operatorname{tr} (n I_m + M'M)$, the remainder being of order $o(n^{-1})$.

Insertion yields

$$2\alpha E\,(\mathrm{tr}\,S)^{-1}\,(M'M)^{-1}\,S\,(M'M)^{-1} - \alpha E\,(\mathrm{tr}\,S)^{-1}\,S\,(M'M)^{-2} - \\ -\alpha E\,(\mathrm{tr}\,S)^{-1}\,(M'M)^{-2}\,S = O + o(n^{-1}).$$

Hence to the order of approximation

$$EL[\,(M'M)^{-1}, T] - EL[\,(M'M)^{-1}, T_\alpha] > \alpha\,[4(n-4) - \alpha]\,E\,(\mathrm{tr}\,S)^{-2}\,(M'M)^{-2} + \\ +2\alpha\left[E\,(\mathrm{tr}\,S)^{-1}\,(M'M)^{-1} - E_1\,(\mathrm{tr}\,S)^{-1}\,(M'M)^{-1}\right] + 4\alpha E_1\,(\mathrm{tr}\,S)^{-2}\,(M'M)^{-1} > O,$$

as $E\,(\mathrm{tr}\,S)^{-1} \geqslant E_1\,(\mathrm{tr}\,S)^{-1}$.

For the auxiliary inequality see Leung (1994, p. 112). □

Appendix 1: a lemma on the matrix Haffian $\nabla\varphi F$

Lemma 4

$$\nabla\varphi F = \varphi\nabla F + \frac{\partial\varphi}{\partial X}F,$$

where φ is a scalar function of the symmetric matrix variable X and F is a matrix function thereof. Further

$$\frac{\partial\varphi}{\partial X} := \tfrac{1}{2}\sum_{ij}\frac{\partial\varphi}{\partial x_{ij}}\left(E_{ij} + E_{ji}\right), \quad \text{where } E_{ij} := e_i e_j'$$

Proof.

$$(\nabla\varphi F)_{ik} = \sum_j d_{ij}\,(\varphi F)_{jk} = \sum_j d_{ij}\varphi f_{jk} = \tfrac{1}{2}\sum_j\left(1 + \delta_{ij}\right)\frac{\partial\varphi f_{jk}}{\partial x_{ij}} = \\ = \frac{\partial\varphi f_{ik}}{\partial x_{ii}} + \tfrac{1}{2}\sum_{j\neq i}\frac{\partial\varphi f_{jk}}{\partial x_{ij}} = \varphi\left(\frac{\partial f_{ik}}{\partial x_{ii}} + \tfrac{1}{2}\sum_{j\neq i}\frac{\partial f_{jk}}{\partial x_{ij}}\right) + \left(\frac{\partial\varphi}{\partial x_{ii}}\right)f_{ik} + \\ + \tfrac{1}{2}\sum_{j\neq i}\frac{\partial\varphi}{\partial x_{ij}}f_{jk} = \varphi\,(\nabla F)_{ik} + \left(\frac{\partial\varphi}{\partial X}\right)_{i.} F_{.k}, \quad \text{hence}$$

$$\nabla\varphi F = \varphi\nabla F + \frac{\partial\varphi}{\partial X}F.$$

Here f_{jk} and $(F)_{jk}$ are the jk^{th} element of F, $F_{i.}$ is the i^{th} row of F and $F_{.j}$ is the k^{th} column of F.

For more details see Neudecker (2000*b*). □

Appendix 2: a lemma on the scalar Haffian $\operatorname{tr}\nabla F_2AF_1$

Lemma 5

$$tr\nabla F_2AF_1 = tr\,(\nabla F_2)AF_1 + tr\,(\nabla F_1')A'F_2',$$

where F_2 and F_1 are functions of the symmetric matrix variable X and A is a constant matrix.

Each F satisfies $F = \sum_k \varphi_k C_k$ or $dF = \sum_l P_l(dX)Q_l'$ with constant C_k, P_l and Q_l.

We consider three cases. The first comprises $F_1 = \varphi C$ and $dF_2 = P(dX)Q'$, the second comprises $F_2 = \varphi C$ and $dF_1 = P(dX)Q'$, the third comprises $dF_1 = P(dX)Q'$ and $dF_2 = R(dX)T'$. The fourth case with $F_1 = \varphi_1 C_1$ and $F_2 = \varphi_2 C_2$ follows easily. Without loss of generality the summation signs were dropped.

Proof.

Case 1. We have $dF_1 = (d\varphi)C$, hence by Lemma 4 $\nabla F_1' = \frac{\partial\varphi}{\partial X}C'$. Further

$$\begin{aligned} d(F_2AF_1) &= (dF_2)AF_1 + F_2AdF_1 \\ &= P(dX)Q'AF_1 + (d\varphi)F_2AC \end{aligned}$$

which implies

$$\begin{aligned} \nabla F_2AF_1 &= \tfrac{1}{2}P'Q'AF_1 + \tfrac{1}{2}(\operatorname{tr}P)Q'AF_1 + \frac{\partial\varphi}{\partial X}F_2AC, \\ \operatorname{tr}\nabla F_2AF_1 &= \tfrac{1}{2}\operatorname{tr}P'Q'AF_1 + \tfrac{1}{2}(\operatorname{tr}P)\operatorname{tr}Q'AF_1 + \operatorname{tr}\frac{\partial\varphi}{\partial X}F_2AC; \\ (\nabla F_2)AF_1 &= \tfrac{1}{2}P'Q'AF_1 + \tfrac{1}{2}(\operatorname{tr}P)Q'AF_1, \\ \operatorname{tr}(\nabla F_2)AF_1 &= \tfrac{1}{2}\operatorname{tr}P'Q'AF_1 + \tfrac{1}{2}(\operatorname{tr}P)\operatorname{tr}Q'AF_1; \\ \left(\nabla F_1'\right)A'F_2' &= \frac{\partial\varphi}{\partial X}C'A'F_2', \\ \operatorname{tr}\left(\nabla F_1'\right)A'F_2' &= \operatorname{tr}\frac{\partial\varphi}{\partial X}C'A'F_2' = \operatorname{tr}\frac{\partial\varphi}{\partial X}F_2AC. \end{aligned}$$

This yields the result.

Case 2. We replace F_1 by F_2', A by A' and F_2 by F_1' in the first result. This leads to

$$\operatorname{tr}\nabla F_1'A'F_2' = \operatorname{tr}(\nabla F_1')A'F_2' + \operatorname{tr}(\nabla F_2)AF_1.$$

Using $\operatorname{tr}\nabla F' = \operatorname{tr}\nabla F$ we get

$$\operatorname{tr}\nabla F_2AF_1 = \operatorname{tr}(\nabla F_1')A'F_2' + \operatorname{tr}(\nabla F_2)AF_1.$$

Case 3. Now $dF_1 = P(dX)Q'$ and $dF_2 = R(dX)T'$. Then

$$\begin{aligned} 2\nabla F_1 &= P'Q' + (\operatorname{tr} P)\, Q' \\ 2\nabla F_2 &= R'T' + (\operatorname{tr} R)\, T' \\ 2\nabla F_1' &= Q'P' + (\operatorname{tr} Q)\, P' \end{aligned}$$

by the Theorem in Neudecker (2000*b*).
Further

$$\begin{aligned} dF_2AF_1 &= (dF_2)AF_1 + F_2AdF_1, \\ &= R(dX)T'AF_1 + F_2AP(dX)Q', \end{aligned}$$

which implies

$$\begin{aligned} 2\nabla F_2AF_1 &= R'T'AF_1 + P'A'F_2'Q' + \\ &\quad + (\operatorname{tr} R)\, T'AF_1 + (\operatorname{tr} F_2AP)\, Q' \\ &= 2(\nabla F_2)AF_1 + P'A'F_2'Q' + (\operatorname{tr} F_2AP)\, Q' \end{aligned}$$

and hence

$$\begin{aligned} 2\operatorname{tr}\nabla F_2AF_1 &= 2\operatorname{tr}\,(\nabla F_2)AF_1 + \operatorname{tr}\,[Q'P' + (\operatorname{tr} Q)\, P']\, A'F_2' \\ &= 2\operatorname{tr}\,(\nabla F_2)AF_1 + 2\operatorname{tr}\left(\nabla F_1'\right)A'F_2'. \end{aligned}$$

□

Note: For an introduction to the scalar Haffian see Neudecker (2000*a*).

5 References

Chow, M. S. (1987). A complete class theorem for estimating a noncentrality parameter. *Ann. Statist.*, 15, 800-4.

Haff, L. R. (1981). Further identities for the Wishart distribution with applications in regression. *Canadian J. Statist.*, 215-24.

Leung, P. L. (1994). An identity for the noncentral Wishart distribution with application. *J. Multivariate Anal.*, 48, 107-14.

Leung, P. L. & Muirhead, R. J. (1987). Estimation of parameter matrices and eigenvalues in MANOVA and canonical correlation analysis. *Ann. Statist.*, 15, 1651-66.

Neudecker, H. (2000*a*). A note on the scalar Haffian. *Qüestiió*, 24(2), 243-9.

Neudecker, H. (2000*b*). A note on the matrix Haffian. *Qüestiió*, 24(3), 419-24.

Perlman, M. D. & Rasmussen, U. A. (1975). Some remarks on estimating a noncentrality parameter. *Comm. Statist.*, 4, 455-68.

Saxena, K. M. L. & Alam, K. (1982). Estimation of the noncentrality parameter of a chi-squared distribution. *Ann. Statist.*, 10, 1012-6.

SORT 28 (2) July-December 2004, 201-214

Some discrete exponential dispersion models: Poisson-Tweedie and Hinde-Demétrio classes

Célestin C. Kokonendji, Simplice Dossou-Gbété*, Clarice G.B. Demétrio**

* *University of Pau-LMA, Pau, France,* ** *University of São Paulo*

Abstract

In this paper we investigate two classes of exponential dispersion models (EDMs) for overdispersed count data with respect to the Poisson distribution. The first is a class of Poisson mixture with positive Tweedie mixing distributions. As an approximation (in terms of unit variance function) of the first, the second is a new class of EDMs characterized by their unit variance functions of the form $\mu + \mu^p$, where p is a real index related to a precise model. These two classes provide some alternatives to the negative binomial distribution ($p = 2$) which is classically used in the framework of regression models for count data when overdispersion results in a lack of fit of the Poisson regression model. Some properties are then studied and the practical usefulness is also discussed.

MSC: 62E15, 62E17, 60E07

Keywords: Negative binomial distribution, overdispersion, Poisson mixture, Tweedie family, unit variance function

1 Introduction

The Poisson distribution is well-known to be the classical distribution for count data, but it has only one parameter and its variance is equal to the mean. Since the index of dispersion (i.e. the variance divided by the mean) of Poisson is one, this makes it inadequate for fitting overdispersed count data (e.g. Castillo and Pérez-Casany, 2004), and raises the question of whether an appropriate two-parameter distribution such as the negative binomial should be used routinely for analysing overdispersed count data. The same problem occurs in the framework of regression models for count data

* *Address for correspondence:* C.C. Kokonendji. Université de Pau et des Pays de l'Adour. Laboratoire de Mathématiques Appliquées. CNRS FRE 2570-Département STID. Avenue de l'Université. 64000 Pau, France. Phone: +33 559 407 145. Fax: +33 559 407 140.

** University of São Paulo - ESALQ, Piracicaba - SP, Brazil.

Received: January 2004

Accepted: September 2004

(McCullagh and Nelder, 1989), when the Poisson distribution does not fit well, and the observed dispersion is greater than that predicted by the standard distribution.

It is well known that negative binomial can be understood as a Poisson mixture with gamma mixing distribution, taking into account the heterogeneity in the population. Hougaard *et al* (1997) have considered a large family of mixture distributions, including the Poisson-inverse Gaussian distribution, to improve significantly the fitness to certain data. We will call the *Poisson-Tweedie* class a completed set of these distributions that we must point out the exact form of its associated "unit variance function" (a term to be made precise). Otherwise, Hinde and Demétrio (1998, page 14) propose for overdispersed count data the use of the unit variance function

$$V_p^{HD}(\mu) = \mu + \mu^p, \quad \mu \in M_p^{HD} \subseteq \mathbb{R}, \tag{1}$$

where $p \in \mathbb{R}$ fixed, which is also an alternative to negative binomial unit variance function obtained with $p = 2$ and includes the strict arcsine distribution with $p = 3$ (Kokonendji and Khoudar, 2004). We here call the *Hinde-Demétrio* class the set of all distributions associated to (1). The aim of this work is to provide a complete identification of both the Poisson-Tweedie and the Hinde-Demétrio classes from their unit variance functions. These classes are sets of two-parameter distributions with an additional index parameter p allowing to identify an appropriate family of these distributions.

In Section 2, we review some basic properties of the general models, called "exponential dispersion models" and, in particular, we present the *Tweedie* class with unit variance function μ^p. In Section 3, we describe the possible Poisson mixture distributions with a Tweedie for obtaining the Poisson-Tweedie class: unit variance function and probabilities are given. In Section 4, we first classify the Hinde-Demétrio class (1) and we then compare it to the Poisson-Tweedie class. We stress that there is no intersection between the Hinde-Demétrio class ($\mu + \mu^p$) and the Tweedie class (μ^p), except for $p = 2$. Section 5 is devoted to concluding remarks and the problem of statistical inference for p to select the adequate model in these classes.

2 Exponential dispersion models

Exponential dispersion models (Jørgensen, 1997) are important in statistical modelling. They have a number of important mathematical properties, which are relevant in practice. They include several well-known families of distributions as special cases, giving a convenient general framework. Generalized linear models (McCullagh and Nelder, 1989) are based on these families of distributions.

Let ν be a σ-finite positive measure on the real line $\mathbb{R}$ (not necessarily a probability) and define the cumulant function K by

$$K(\theta) = \ln \int_{\mathbb{R}} \exp(\theta x)\nu(dx)$$

on its (canonical parameter) domain $\Theta = \{\theta \in \mathbb{R} : K(\theta) < \infty\}$. Assume that both ν and Θ are not degenerate (i.e., ν is not concentrated at one point and the interior of Θ is not empty), then the set of the probability measures $P(\theta;\nu)(dx) = \exp\{\theta x - K(\theta)\}\nu(dx)$ defined for all θ in $\Theta = \Theta(\nu)$ represents a *natural exponential family* (NEF) generated by ν and denoted $F = F(\nu) = \{P(\theta;\nu); \theta \in \Theta(\nu)\}$; see Chapter 54 of Kotz *et al* (2000). Given a NEF, we define the set Λ of reals $\lambda > 0$ such that $\lambda K(\theta)$ is also the cumulant function for some measure ν_λ. For fixed $\lambda \in \Lambda$, the NEF $F_\lambda = F(\nu_\lambda)$ generated by ν_λ is then $\exp\{\theta x - \lambda K(\theta)\}\nu_\lambda(dx)$, for $\theta \in \Theta$. This family of distributions, denoted $\mathcal{ED}(\theta,\lambda)$ for $(\theta,\lambda) \in \Theta \times \Lambda$, is called the *exponential dispersion model* (EDM) generated by ν (or ν_λ for improper notation); and λ can be called the dispersion parameter. Its density or mass function with respect to some measure η can be written as

$$C(x;\lambda)\exp\{\theta x - \lambda K(\theta)\}, \quad x \in S \subseteq \mathbb{R}, \tag{2}$$

where $\nu_\lambda(dx) = C(x;\lambda)\eta(dx)$. Note here that $\mathcal{ED}(\theta,\lambda)$ defined by (2) is the additive version of EDM. The reproductive version of $X \sim \mathcal{ED}(\theta,\lambda)$ is given by $Z = X/\lambda$. However, additive EDMs turn out to be important for discrete data because many usefull families of discrete distributions have this form. Any EDM satisfies $\mathcal{ED}(\theta,\lambda_1) * \mathcal{ED}(\theta,\lambda_2) = \mathcal{ED}(\theta,\lambda_1+\lambda_2)$, so the family is closed under convolution and $\{1,2,\cdots\} \subseteq \Lambda$. Also the model is infinitely divisible if and only if $\Lambda = (0,\infty)$.

In the interior of Θ, denoted intΘ, the cumulant function $\theta \mapsto K(\theta)$ is strictly convex. Then the expectation and variance of $X \sim \mathcal{ED}(\theta,\lambda)$ are

$$\mathbb{E}(X) = \lambda K'(\theta) \quad \text{and} \quad \mathrm{Var}(X) = \lambda K''(\theta), \tag{3}$$

where $K'(\theta)$ and $K''(\theta)$ are, respectively, the first and second derivatives of K at the point θ. From (3) with $\lambda = 1$, the characterizing function V defined on the domain $M = K'(\mathrm{int}\Theta)$ such that

$$K''(\theta) = V\{K'(\theta)\}$$

is called *unit variance function*. We also have $V(\mu) = 1/\psi'(\mu)$, for $\mu \in M$, where $\psi = (K')^{-1}$ is the inverse function of K'. Note that M depends only on the family $F = \{\mathcal{ED}(\theta,1) : \theta \in \Theta\}$, and not on the choice of the generating measure ν of F. If $M = \Omega$, where Ω denotes the interior of the convex hull of the support S of F, the family F is said to be *steep*. From here to the end, an EDM is always assumed to be steep. The role of the unit variance function in data fitting should be to identify an appropriate EDM of distributions, if any. The reparametrization by unit mean $\mu = K'(\theta) = \mu(\theta)$ allows us to write the EDM as follows: $\{\mathcal{ED}(\mu(\theta),\lambda); \mu(\theta) \in M, \lambda \in \Lambda\} \equiv EDM(\mu,\lambda)$. It is sometimes considered the reparametrization of the EDM by the mean $m = \mathbb{E}(X) = \lambda K'(\theta) = m(\lambda,\theta)$ instead of the unit mean $\mu = \mu(\theta)$. From (3), the unit variance function V leads to the variance $V_\lambda = \mathrm{Var}(X)$ of $X \sim \mathcal{ED}(\theta,\lambda)$ in terms of m, called *variance function* and expressed as follows: $V_\lambda(m) = \lambda V(m/\lambda)$, for all $m \in \lambda M$. For discrete overdispersed EDM compared to the Poisson distribution we have

$$V(\mu) > \mu, \quad \mu > 0, \tag{4}$$

where $V(\mu) = \mu$ is the unit variance function of the Poisson model (e.g. Jourdan and Kokonendji, 2002).

A complete description of the EDMs with power unit variance functions

$$V_p^T(\mu) = \mu^p, \quad p \in (-\infty, 0] \cup [1, \infty), \tag{5}$$

is given by Jørgensen (1997) where, for $p \to \infty$ the corresponding unit variance function takes the exponential form $V_\infty^T(\mu) = \exp(\beta\mu)\,, \beta \neq 0$. This class, called the *Tweedie class*, was introduced by Tweedie (1984). It is also convenient to introduce the index parameter α of stable distribution, connected to p by the following relation:

$$(p-1)(1-\alpha) = 1. \tag{6}$$

According to the above notations, we can denote by $\mathcal{T}_p(\theta, \lambda)$ any distribution of this class where $\lambda \in (0, \infty) = \Lambda$ for all p of (5), $\mu \in M_p = K'(\text{int}\Theta_p)$ and $\theta \in \Theta_p$ with

$$\Theta_p = \begin{cases} \mathbb{R} & \text{for} \quad p = 0, 1 \\ [0, \infty) & \text{for} \quad p < 0 \ \text{ or } 0 < p < 1 \\ (-\infty, 0) & \text{for} \quad 1 < p \leq 2 \ \text{ or } p \to \infty \\ (-\infty, 0] & \text{for} \quad 2 < p < \infty. \end{cases} \tag{7}$$

Thus, for $s \in \Theta_p - \theta$ and $X \sim \mathcal{T}_p(\theta, \lambda)$, the Laplace transform $\mathbb{E}(e^{sX})$ is

$$G_p(s; \theta, \lambda) = \begin{cases} \exp\left\{\dfrac{\lambda[(1-p)\theta]^\alpha}{(2-p)}[(1+s/\theta)^\alpha - 1]\right\} & \text{for} \quad p \neq 1, 2 \\ (1+s/\theta)^{-\lambda} & \text{for} \quad p = 2 \\ \exp\{\lambda e^\theta (e^s - 1)\} & \text{for} \quad p = 1. \end{cases} \tag{8}$$

As shown in Table 1, the Tweedie class $T = \{TM_p(\mu, \lambda); p \in \mathbb{R}\}$ includes several well-known families of distributions amongst which one may be the inverse-Gaussian model $TM_3(\mu, \lambda)$ and the noncentral gamma model $TM_{3/2}(\mu, \lambda)$ of zero shape (respectively, a special case of positive stable and compound Poisson families). The compound Poisson $(1 < p < 2)$ is also called Poisson-gamma; it can be represented as the Poisson random sum of independent gamma random variables (and it has a mass at zero but otherwise has a continuous positive distribution). Observe, however, that the extreme stable distributions $(p < 0)$ are not steep and only the Poisson distribution $(p = 1)$ is discrete.

Table 1: *Summary of Tweedie EDMs (Jørgensen, 1997).*

Distribution	p	α	M	S
Extreme stable	$p < 0$	$1 < \alpha < 2$	$(0, \infty)$	$\mathbb{R}$
Normal	$p = 0$	$\alpha = 2$	$\mathbb{R}$	$\mathbb{R}$
[Do not exist]	$0 < p < 1$	$2 < \alpha < \infty$		
Poisson	$p = 1$	$\alpha \to -\infty$	$(0, \infty)$	$\mathbb{N}$
Compound Poisson	$1 < p < 2$	$\alpha < 0$	$(0, \infty)$	$(0, \infty)$
Gamma	$p = 2$	$\alpha = 0$	$(0, \infty)$	$(0, \infty)$
Positive stable	$p > 2$	$0 < \alpha < 1$	$(0, \infty)$	$[0, \infty)$
Extreme stable	$p \to \infty$	$\alpha = 1$	$\mathbb{R}$	$\mathbb{R}$

3 Poisson-Tweedie EDMs

Let X be a non-negative random variable following $\mathcal{T}_p(\theta, \lambda)$. If a discrete random variable Y is such that the conditional distribution of Y given X is Poisson with mean X, then the EDM generated by the distribution of Y is of the Poisson-Tweedie class. We can also use the following notations $\mathcal{PT}_p(\theta, \lambda)$ to denote the distribution of Y and $PTM_p(\mu, \lambda)$ for the corresponding EDM. Hence for $p \geq 1$, the individual probabilities of $Y \sim \mathcal{PT}_p(\theta, \lambda)$ are

$$\Pr(Y = y) = \int_0^\infty \frac{e^{-x}x^y}{y!} \mathcal{T}_p(\theta, \lambda)(dx), \quad y = 0, 1, \cdots. \tag{9}$$

Proposition 1 *(Hougaard et al., 1997) Let $Y \sim \mathcal{PT}_p(\theta, \lambda)$ defined by (9), where $\theta \in \Theta_p$ given by (7) and $\lambda > 0$ for fixed $p \geq 1$ or $\alpha \in [-\infty, 1)$ from (6). We have the following properties: (i) If $Y_1, \cdots, Y_n$ are independent, with $Y_i \sim \mathcal{PT}_p(\theta, \lambda_i)$, then $Y_1 + \cdots + Y_n$ follows $\mathcal{PT}_p(\theta, \lambda_1 + \cdots + \lambda_n)$. The distribution $\mathcal{PT}_p(\theta, \lambda)$ is infinitely divisible. (ii) The distribution $\mathcal{PT}_p(\theta, \lambda)$ is unimodal for $p \geq 2$. (iii) The Laplace transform of $Y \sim \mathcal{PT}_p(\theta, \lambda)$ is*

$$\mathbb{E}(e^{\omega Y}) = \begin{cases} \exp\left\{\dfrac{\lambda}{2-p}[\{(1-p)(e^\omega - 1 + \theta)\}^\alpha - \{(1-p)\theta\}^\alpha]\right\} & \text{for} \quad p \neq 1, 2 \\ [(e^\omega - 1 + \theta)/\theta]^{-\lambda} & \text{for} \quad p = 2 \\ \exp\left\{\lambda[\exp(e^\omega - 1 + \theta) - e^\theta]\right\} & \text{for} \quad p = 1, \end{cases} \tag{10}$$

for $\omega \in \Theta_p - \theta$. For $p = 1$, it is a Neyman type A distribution; for $p = 2$, it is a negative binomial distribution; and, for $p = 3$, it is the Sichel or Poisson-inverse Gaussian distribution (e.g. Willmot, 1987).

Proposition 2 *With the asumptions of Proposition 1, the unit variance function of the model* $PTM_p(\mu, \lambda)$ *generated by* $Y \sim \mathcal{PT}_p(\theta, 1)$ *is exactly*

$$V_p^{PT}(\mu) = \mu + \mu^p \exp\{(2-p)\Phi_p(\mu)\}, \quad \mu > 0, \tag{11}$$

where $\Phi_p(\mu)$*, generally implicit, denotes the inverse of the increasing function* $\omega \mapsto d\{\ln \mathbb{E}(e^{\omega Y})\}/d\omega$.

Proof. Let $K(\omega) = \ln \mathbb{E}(e^{\omega Y})$ for $Y \sim \mathcal{PT}_p(\theta, 1)$. From Proposition 1 *(iii)* with $\lambda = 1$ and using (6) to simplify, the first derivative of $K(\omega)$ is

$$\mu = K'(\omega) = \begin{cases} e^{\omega}[(1-p)(e^{\omega} - 1 + \theta)]^{\alpha-1} & \text{for} \quad p \neq 1, 2 \\ -e^{\omega}(e^{\omega} - 1 + \theta)^{-1} & \text{for} \quad p = 2 \\ e^{\omega}\exp\{e^{\omega} - 1 + \theta\} & \text{for} \quad p = 1, \end{cases}$$

and the second derivative of $K(\omega)$ may be expressed as follows:

$$V_p^{PT}(\mu) = K''(\omega) = \begin{cases} K'(\omega) + e^{2\omega}[(1-p)(e^{\omega} - 1 + \theta)]^{\alpha-2} & \text{for} \quad p \neq 1, 2 \\ K'(\omega) + [K'(\omega)]^2 & \text{for} \quad p = 2 \\ K'(\omega) + e^{\omega}K'(\omega) & \text{for} \quad p = 1. \end{cases}$$

For $p \neq 1, 2$, we can also write $K''(\omega) = K'(\omega) + e^{2\omega}[K'(\omega)/e^{\omega}]^p$ and the expression given in (11) is easily obtained. □

The Poisson-Tweedie EDMs are summarized in Table 2, that we can divide in three parts with respect to p: $1 < p < 2$, $2 < p < \infty$ and $p \in \{1, 2, \infty\}$.

***Table 2**: Summary of Poisson-Tweedie EDMs.*

Distribution	p	α	M	S
[Do not define]	$p < 1$	$1 < \alpha < \infty$		
Neyman type A	$p = 1$	$\alpha \to -\infty$	$(0, \infty)$	$\mathbb{N}$
Poisson-compound Poisson	$1 < p < 2$	$\alpha < 0$	$(0, \infty)$	$\mathbb{N}$
Negative binomial	$p = 2$	$\alpha = 0$	$(0, \infty)$	$\mathbb{N}$
Poisson-positive stable	$p > 2$	$0 < \alpha < 1$	$(0, \infty)$	$\mathbb{N}$
Poisson	$p \to \infty$	$\alpha = 1$	$(0, \infty)$	$\mathbb{N}$

For $p = 1$ or $\alpha \to -\infty$ which is not studied by Hougaard *et al.* (1997), we can refer to Johnson *et al.* (1992; pages 368-) to obtain some properties on the Neyman type A distribution, which is therefore both a Poisson mixture of Poisson distributions, and also a Poisson-stopped sum of Poisson distributions.

Note that we consider in this paper only the strictly mixed Poisson distributions ($p \geq 1$). In fact, it is not possible to mix a Poisson with a Tweedie distribution $\mathcal{T}_p(\theta, \lambda)$ for $p \leq 0$ because it can be negative but (9) can be seen as a purely formal operation. For $p = 0$ we refer to Kemp and Kemp (1966) who show that, if X follows a normal distribution $\mathcal{T}_0(\theta, \lambda)$ with mean $\mu = \mu(\theta)$ and standard deviation $\sigma = \sigma(\lambda)$ such that $\mu \geq \sigma^2$, the corresponding mixed Poisson distribution is the Hermite distribution (Johnson *et al.*, 1992; pages 357-364).

To conclude this section, we explicit the probability mass functions (2) of all the Poisson-Tweedie EDMs generated by any distribution of $Y \sim \mathcal{PT}_p(\theta_0, \lambda)$ for fixed $p \in [1, \infty)$ and $\theta_0 \in \bar{\Theta}_p$ the closure of Θ_p given in (7). Indeed, one has

$$C_{p,\theta_0}(y; \lambda) \exp\{\omega y - \lambda K_{p,\theta_0}(\omega)\}, \quad y = 0, 1, 2, \cdots, \tag{12}$$

where $\omega \in \Theta_p - \theta_0$ is the canonical parameter and $\lambda > 0$ is the dispersion parameter such that, respectively by (9) and (10),

$$C_{p,\theta_0}(y; \lambda) = \frac{1}{y!}\mathbb{E}(e^{-X}X^y) = \frac{1}{y!}\frac{\partial^y G_p(s; \theta_0, \lambda)}{\partial s^y}|_{s=-1} \text{ for } X \sim \mathcal{T}_p(\theta_0, \lambda) \tag{13}$$

and $K_{p,\theta_0}(\omega) = \lambda^{-1} \ln \mathbb{E}(e^{\omega Y})$ for $Y \sim \mathcal{PT}_p(\theta_0, \lambda)$. Note that, in practice, we can use $\theta_0 = 0$ which is here defined for all $p \in [1, \infty)$ with the convenience: $0^\lambda = 1$, for any $\lambda > 0$ (only for $p = 2$). To clarify completely (12), we must point out $C_{p,\theta_0}(y; \lambda)$ by the following proposition.

Proposition 3 *Let $p \in [1, \infty)$ be fixed and let $\theta_0 \in \bar{\Theta}_p$ given by (7). Then, for all $\lambda > 0$ and $y \in \mathbb{N}$, the expression of $C_{p,\theta_0}(y; \lambda)$ in (12) is,*
- for $p = 1$ or $\alpha \to -\infty$,

$$C_{1,\theta_0}(y; \lambda) = \begin{cases} \exp\{\lambda e^{\theta_0 - 1}(1 - e)\} & \text{for } y = 0 \\ \dfrac{1}{y!} C_{1,\theta_0}(0; \lambda) \sum\limits_{k=1}^{y} a_{y,k} (\lambda e^{\theta_0 - 1})^k & \text{for } y = 1, 2, \cdots, \end{cases}$$

with $a_{y,y} = a_{y,1} = 1$ and $a_{y,k} = a_{y-1,k-1} + k a_{y-1,k}$;

– for $p = 2$ or $\alpha = 0$, $C_{2,\theta_0}(y; \lambda) = \left(\dfrac{\theta_0}{\theta_0 - 1}\right)^\lambda \dfrac{\Gamma(y + \lambda)}{y!\Gamma(\lambda)} \left(\dfrac{1}{1 - \theta_0}\right)^y$;

– for $p \in (1, 2) \cup (2, \infty)$ or $\alpha \in (-\infty, 0) \cup (0, 1)$,

$$C_{p,\theta_0}(y; \lambda) = \begin{cases} \exp\left\{\dfrac{\lambda(\alpha - 1)}{\alpha}[\{(1 - \theta_0)/(1 - \alpha)\}^\alpha - \{(-\theta_0)/(1 - \alpha)\}^\alpha]\right\} & \text{for } y = 0 \\ \dfrac{1}{y!} C_{p,\theta_0}(0; \lambda) \sum\limits_{k=1}^{y} a_{y,k}(\alpha) \lambda^k \{(1 - \theta_0)/(1 - \alpha)\}^{k\alpha - y} & \text{for} y = 1, 2, \cdots, \end{cases}$$

with $a_{y,y}(\alpha) = 1$, $a_{y,1}(\alpha) = \Gamma(y - \alpha)/[(1 - \alpha)^{y-2}\Gamma(2 - \alpha)]$ and $a_{y,k}(\alpha) = a_{y-1,k-1}(\alpha) + [(y - 1 - k\alpha)/(1 - \alpha)]a_{y-1,k}(\alpha)$.

Proof. Following Hougaard *et al.* (1997), we show only for $p = 1$ or $\alpha \to -\infty$. From (13), it suffices to check the recursive formula of derivatives of Laplace transform (8), which is $\partial^y G_1(s; \theta_0, \lambda)/\partial s^y = G_1(s; \theta_0, \lambda) \sum_{k=1}^{y} a_{y,k}(\lambda e^{\theta_0 + s})^k$. □

4 Hinde-Demétrio EDMs

We now characterize the Hinde-Demétrio class which is the set of EDMs with unit variance function of the "simple" form (1) and, then, we compare it to the Poisson-Tweedie class (11).

Theorem 4 *Let $p > 1$. Then there exists a NEF $F_{p,1}$ such that $M_{F_{p,1}} = (0, \infty)$ and $V_{F_{p,1}}(\mu) = \mu + \mu^p$. Furthermore, the NEF $F_{p,1}$ is infinitely divisible (with bounded Lévy measure). More precisely, denote $a = 1/(p-1)$ similarly to the interchangeable relation (6) and consider the positive measure*

$$\nu_p = \sum_{k=0}^{\infty} \frac{a(a+1)\cdots(a+k-1)}{k!} \frac{1}{1+k(p-1)} \delta_{1+k(p-1)}.$$

Then, for all $\lambda > 0$, $\nu_{p,\lambda} = \exp\{\lambda \nu_p\} = \sum_{n=0}^{\infty} \frac{\lambda^n}{n!} \nu_p^{*n}$ generates a NEF $F_{p,\lambda}$ with variance function

$$V_{F_{p,\lambda}}(m) = m + \lambda^{1-p} m^p \tag{14}$$

and $F_{p,\lambda}$ is concentrated on the additive semigroup $\mathbb{N} + p\mathbb{N}$.

Proof. It is standard by checking that we have exactly

$$K'_{\nu_{p,1}}(\theta) = \frac{e^\theta}{[1 - e^{\theta(p-1)}]^a} = \sum_{k=0}^{\infty} \frac{a(a+1)\cdots(a+k-1)}{k!} e^{\theta[1+k(p-1)]}.$$

□

For $p > 1$, we also denote by $\mathcal{HD}_p(\theta, \lambda)$ any distribution of the EDM $HDM_p(\mu, \lambda)$ corresponding to (1) or (14) with $\lambda > 0$ and $\theta \in \Theta_p^* \subseteq \mathbb{R}$. As a probabilistic interpretation of $Y \sim \mathcal{HD}_p(\theta, \lambda)$, let X be a random variable associated to $\nu_p \equiv \nu_{p,\theta}$ of Theorem 4 (up to normalizing constant) and, let N_t be a standard Poisson process on the interval $(0, t]$ ($N_0 = 0$) with rate λ (i.e. $N_t \sim \mathcal{P}(\lambda t)$) and supposed to be independent of X. From the Laplace transform of

$$Y_t = \sum_{i=1}^{N_t} X_i = X_1 + \cdots + X_{N_t},$$

where the X_i are independent and identically distributed as X, it is easy to see that $Y = Y_1$ by fixing the time to $t = 1$.

Theorem 5 *Let $p \in \mathbb{R}$. The function (1): $\mu \mapsto V_p^{HD}(\mu) = \mu + \mu^p$, defined on a suitable domain M_p^{HD} corresponds to a unit variance function of discrete (steep) EDM when*

$$p \in \{0\} \cup [1, \infty),$$

with $M_0^{HD} = (-1, \infty)$ and $M_p^{HD} = (0, \infty)$ for $p \geq 1$; and the domain Θ_p of the canonical parameter is given by (7). Moreover, if $p = 0$ the model $HDM_0(\mu, \lambda)$ is a positive-translated Poisson; if $p = 1$ the model $HDM_1(\mu, \lambda)$ is a scaled Poisson; if $p = 2$

the model $HDM_2(\mu, \lambda)$ is negative binomial; if $p = 3$ the model $HDM_3(\mu, \lambda)$ is strict arcsine (Kokonendji and Khoudar, 2004); and, if $p \neq 0, 1, 2, 3$ the model $HDM_p(\mu, \lambda)$ is deduced from Theorem 4.

Before giving the proof, let us observe that, from (4), the Hinde-Demétrio class (1) is the set of overdispersed EDMs compared with the Poisson distribution, as well as the Poisson-Tweedie class (11) (see also Feller, 1943). The next proposition provides comparable behaviours of these two classes. Indeed, only negative binomial model $HDM_2(\mu, \lambda)$ is interpreted as $PTM_2(\mu, \lambda)$; and, for fixed $p \geq 1$ and $\lambda > 0$, each $HDM_p(\mu, \lambda)$ can be "approximated" by $PTM_p(\mu, \lambda)$ for large μ. This approximation must be understand in terms of their unit variance functions.

Proposition 6 *Let $HD = \{HDM_p(\mu, \lambda); p \in \mathbb{R}\}$ be the Hinde-Demétrio class (1) and $PT = \{PTM_p(\mu, \lambda); p \in \mathbb{R}\}$ the Poisson-Tweedie class (11). Then: (i) $HD \cap PT = \{HDM_2(\mu, \lambda) = PTM_2(\mu, \lambda)\}$; (ii) For fixed $p \geq 1$, $V_p^{PT}(\mu) \sim V_p^{HD}(\mu)$ as $\mu \to \infty$.*

For the proof of Theorem 5 we need the two following lemmas. The first is an "impossibility criterion" to exclude case $0 < p < 1$, and the second is related to the steepness.

Lemma 7 *There are no EDM with $M = (0, \infty)$ and unit variance function $V(\mu) \sim \mu^\gamma$ as $\mu \to 0$ for $\gamma \in (0, 1)$.*

Proof. If $V(\mu) \sim \mu^\gamma$ as $\mu \to 0$, then $\theta = \psi(\mu) = \theta_0 + \int_0^\mu (V(t))^{-1} dt$ is left-bounded. Now, $V(\mu) \to 0$ as $\mu \to 0$ implies that the generator ν is concentrated on $[0, \infty)$ (e.g. Letac and Mora, 1990). Hence, the domain $\Theta(\nu)$ is not left-bounded, which yields a contradiction. □

Lemma 8 *(Jørgensen, 1997; page 58) Let $F = \{\mathcal{ED}(\theta, 1); \theta \in \Theta\}$ be a NEF with variance function V on M and support S. If $\inf S = 0$, then: (i) $\inf M = 0$; (ii) $\lim_{\mu \to 0} V(\mu) = 0$; (iii) $\lim_{\mu \to 0} V(\mu)/\mu = c$, where $c = \inf\{S \setminus \{0\}\}$.*

Note that $c = 0$ for continuous distributions and $c = 1$ for discrete integer-valued distributions.

Proof of Theorem 5: Since V_p^{HD} must be an analytic positive function on the domain $M_p^{HD} = (a, \infty)$, we have that V_p^{HD} has no zeroes in (a, ∞) and $V_p^{HD}(a) = 0$ [this is a consequence of Theorem 3.1 of Letac and Mora, 1990]. Thus, we have

$$M_p^{HD} = \begin{cases} (0, \infty) & \text{for} \quad p \neq 0 \\ (-1, \infty) & \text{for} \quad p = 0. \end{cases} \tag{15}$$

Solving $\psi'(\mu) = 1/V_p^{HD}(\mu) = 1/(\mu + \mu^p)$ and ignoring the arbitrary constants in the solutions,

$$\psi(\mu) = \begin{cases} \ln(\mu) - (p-1)^{-1}\ln(1+\mu^{p-1}) & \text{for } \; p \neq 0,1 \\ \ln\sqrt{\mu} & \text{for } \; p = 1 \\ \ln(1+\mu) & \text{for } \; p = 0. \end{cases} \tag{16}$$

We now examine the different situations of $p \in \mathbb{R}$ in (1) from (15).

– Consider case $p \in \{0\} \cup [1,\infty)$. Let $\theta = \psi(\mu)$ be the canonical link function given by (16), then we find first $\mu = K'(\theta)$ and, hence,

$$K(\theta) = \begin{cases} e^{\theta} - \theta & \text{for } \; p = 0 \\ e^{2\theta}/2 & \text{for } \; p = 1 \\ -\ln(1-e^{\theta}) & \text{for } \; p = 2 \\ \arcsin e^{\theta} & \text{for } \; p = 3 \\ \sum_{k=0}^{\infty} \dfrac{\Gamma[k+1/(p-1)]}{k!\Gamma[1/(p-1)]} \dfrac{\exp\{\theta[1+k(p-1)]\}}{1+k(p-1)} & \text{for } \; p \neq 0,1,2,3, \end{cases}$$

for $\theta \in \Theta$, where the interior of Θ is obtained by using (15) and (16):

$$\text{int}\Theta = \begin{cases} \mathbb{R} & \text{for } \; p = 0,1 \\ (0,\infty) & \text{for } \; p < 0 \text{ or } 0 < p < 1 \\ (-\infty,0) & \text{for } \; 1 < p < \infty. \end{cases} \tag{17}$$

Since $K(\theta)$ is analytic, the domain Θ defined from its interior (17) coincides to Θ_p given by (7). Thus, for each $p \in \{0\} \cup [1,\infty)$, we define a discrete generator of the corresponding (steep) EDM with unit variance function (1).

– Case $0 < p < 1$ is excluded by Lemma 7.

– Finally, let us exclude case $p < 0$ by Lemma 8. Indeed, it suffices to observe that $M_p^{HD} = (0,\infty)$ from (15) and $\lim_{\mu\to 0} V_p^{HD}(\mu)/\mu = \lim_{\mu\to 0}(1+\mu^{p-1}) = \infty$. The proof of Theorem 5 is now complete. □

5 Concluding remarks

Here we have two classes of two parameter distributions which could be used as models for analysing overdispersed count data. It was showed that both are EDMs with general unit variance functions indexed by a third parameter p. A common member of both families is the negative binomial family when $p = 2$. However, the probability mass functions (2) of $HDM_p(\mu,\lambda)$ are generally not easy to calculate (except for $p \in \{0,1,2,3\}$) whereas for $PTM_p(\mu,\lambda)$ are given explicitly by (12). When μ is large, Proposition 6 (*ii*) allows the use of the Poisson-Tweedie model as well as the Hinde-Demétrio model since the variance functions are equivalent.

For models with covariates, let Y be a count response variable and let $\mathbf{x}$ be an associated $d \times 1$ vector of covariates with a vector β of unknown regression coefficients, the relation between the mean $m = \mathbb{E}(Y) = m(\mathbf{x};\beta)$ of the distribution and the linear part

$\mathbf{x}^T\boldsymbol{\beta}$ being made through a link function (McCullagh and Nelder, 1989). For both EDMs the cumulant functions are given in explicit form. This allows to compute deviances and then to use maximum likelihood method for the estimation of the parameters.

When the Poisson-Tweedie models $PTM_p(\mu, \lambda)$ or the Hinde-Demétrio models $HDM_p(\mu, \lambda)$ are used, one of the problems for statistical inference is the index parameter p of the adequate distribution. A profile estimate of p is recommended in the general situation; see Hougaard *et al.* (1997) for Poisson-Tweedie mixture. In Hinde-Demétrio models, we can start by the moment estimate of p. Indeed, if $\underline{y} = (y_1, \cdots, y_n)$ is an n-independent identically distributed observation from $\mathcal{HD}_p(\mu, \lambda)$ such that the overdispersion condition $s^2/\bar{y} > 1$ is satisfied, where $\bar{y}$ and s^2 are, respectively, the sample mean and the sample variance from $\underline{y}$. From (14) and when λ is fixed or known, we easily have by the moment method

$$p^* = \ln[(s^2 - \bar{y})/\lambda]/\ln[\bar{y}/\lambda] = p^*(\lambda). \tag{18}$$

With respect to the unit variance function (1), we can take $\lambda = 1$ in (18) because one stays in the same EDM; after, one can estimate λ in the corresponding $HDM_{p^*}(\mu, \lambda)$.

In order to illustrate (18), we analyze two data sets of Table 3 given by Kokonendji and Khoudar (2004; Table 5.1 and Table 5.2); one of which (Table 5.1) has been revisited by several authors (e.g. Willmot, 1987). This type of data is frequent in marketing, insurance, biometry and financial problems. The data correspond to the number of automobile insurance claims per policy in Germany over 1960 among $n = 23589$ and in Central African Republic over 1984 among $n = 10000$. Both data sets show overdipersion as can be seen in Table 4 with $s^2/\bar{y} \simeq 1.14$. For these data some overdispersed models have been used in the literature among others the negative binomial ($HDM_2 = PTM_2$), strict arcsine (HDM_3) and Poisson-inverse Gaussian (PTM_3) models. It was used a Pearson's chi-square goodness-of-fit statistic and, as observed by several authors, no single probability law seems to emerge as providing "the" best fit.

Table 3: *The two data sets analysed by Kokonendji and Khoudar (2004; Table 5.1 and Table 5.2).*

Data set 1		Data set 2	
No. of claims	No. of policies	No. of claims	No. of policies
0	20592	0	6984
1	2651	1	2452
2	297	2	433
3	41	3	100
4	7	4	26
5	0	5+	5
6	1		

From Table 3 we see that using expression (18) with $\lambda = 1$ we find p^* equal to 2 for data set 1 and 3 for data set 2, leading, respectively, to a negative binomial ($HDM_2 = PTM_2$)

and to a strict arcsine (HDM_3) as models of the Hinde-Demétrio class. As Kokonendji and Khoudar (2004) or Willmot (1987) have shown, the best fit for data set 1 is Poisson-inverse Gaussian model (PTM_3) of the Poisson-Tweedie class with $\chi^2 = 0.48$ for 2 degrees of freedom. For data set 2 the value of χ^2 for the strict arcsine model (HDM_3) is bigger than the value of $\chi^2_2(95\%) = 5.99$, meaning that it does not fit; however, it is the best fit among the models of the Hinde-Demétrio class.

Table 4: *Summary statistics for two data sets of Table 3 [the χ^2 values correspond to the negative binomial and strict arcsine models (Kokonendji and Khoudar, 2004)].*

Data sets	$\bar{y}$	s^2	$s^2/\bar{y}$	p^*	χ^2 values	df
1	0.14	0.16	1.14	1.99	3.12	2
2	0.37	0.42	1.14	3.01	15.64	2

Despite the inefficient estimation of parameter p, in Table 3, we have a good way to obtain the adequate model in the Hinde-Demétrio class or even to think to the Poisson-Tweedie class when $p \simeq 2$. However, inferential techniques are not yet as well developed routinely for these two classes of EDMs. But it should be always interesting to handle simultaneously or separately these two classes, instead for using particular distributions.

Acknowledgments We were partially supported by grants from CNPq, FAPESP and CCInt/USP. The authors would like to thank the Editors and a referee for their helpful comments and suggestions.

References

Castillo, J. and Pérez-Casany, M. (2004). Overdispersed and underdispersed Poisson generalizations. *J. Statist. Plann. Inference*, to appear.

Feller, W. (1943). On a general class of contagious distributions. *Ann. Math. Statist.*, 14 (4), 389-400.

Hinde, J. and Demétrio, C. G. B. (1998). *Overdispersion: Models and Estimation.* São Paulo: ABE.

Hougaard, P., Lee, M-L. T. and Whitmore, G. A. (1997). Analysis of overdispersed count data by mixtures of Poisson variables and Poisson processes, *Biometrics*, 53, 1225-1238.

Johnson, N. L., Kotz, S. and Kemp, A. W. (1992). *Univariate Discrete Distributions, 2nd ed.* New York: John Wiley & Sons.

Jørgensen, B. (1997). *The Theory of Dispersion Models.* London: Chapman & Hall.

Jourdan, A. and Kokonendji, C. C. (2002). Surdispersion et modèle binomial négatif généralisé. *Rev. Statistique Appliquée*, L (3), 73-86.

Kemp, A. W. and Kemp, C. D. (1966). An alternative derivation of the Hermite distribution. *Biometrika*, 53, 627-628.

Kokonendji, C. C. and Khoudar, M. (2004). On strict arcsine distribution. *Commun. Statist.-Theory Meth.*, 33 (5), 993-1006.

Kotz, S., Balakrishnan, N. and Johnson, N. L. (2000). *Continuous Multivariate Distributions, vol. 1: Models and Applications*, 2nd ed. New York: Wiley.

Letac, G. and Mora, M. (1990). Natural real exponential families with cubic variance functions. *Ann. Statist.*, 18, 1-37.

McCullagh, P. and Nelder, J. A. (1989). *Generalized Linear Models*, 2nd ed. London: Chapman & Hall.

Tweedie, M. C. K. (1984). An index which distinguishes between some important exponential families. In *Statistics: Applications and new directions. Proceedings of the Indian Statistical Institute Golden Jubilee International Conference* (eds. J.K. Ghosh and J. Roy), 579-604. Indian Statistical Institute, Calcutta.

Willmot, G. E. (1987). The Poisson-inverse Gaussian distribution as an alternative to the negative binomial. *Scandinavian Actuarial J.*, 2, 113-127.

SORT 28 (2) July-December 2004, 215-230

A comparative study of small area estimators⋆

Laureano Santamaría*, Domingo Morales*, Isabel Molina**

* *Universidad Miguel Hernández de Elche,* ** *Universidad Carlos III de Madrid*

Abstract

It is known that direct-survey estimators of small area parameters, calculated with the data from the given small area, often present large mean squared errors because of small sample sizes in the small areas. Model–based estimators borrow strength from other related areas to avoid this problem. How small should domain sample sizes be to recommend the use of model-based estimators? How robust small area estimators are with respect to the rate *sample size/number of domains*?
To give answers or recommendations about the questions above, a Monte Carlo simulation experiment is carried out. In this simulation study, model-based estimators for small areas are compared with some standard design-based estimators. The simulation study starts with the construction of an artificial population data file, imitating a census file of an Statistical Office. A stratified random design is used to draw samples from the artificial population. Small area estimators of the mean of a continuous variable are calculated for all small areas and compared by using different performance measures. The evolution of this performance measures is studied when increasing the number of small areas, which means to decrease their sizes.

MSC: 62D05, 62F10, 62J02

Keywords: small area estimation, *eblup* estimators, sampling designs, mixed linear models

1 Introduction

The problem of small area estimation arises when samples are drawn from (large) populations, but estimates calculated using sample data are required for smaller domains, within which sample data is not enough to provide reliable direct-survey estimators.

⋆ Supported by the grants BMF2003-04820 and GV04B-670.

* Centro de Investigación Operativa, Universidad Miguel Hernández de Elche. Avda. de la Universidad, s/n, 03202-Elche, Spain.

** Departamento de Estadística, Universidad Carlos III de Madrid. C/Madrid, 126, 28903-Getafe, Madrid, Spain.

Received: March 2004
Accepted: July 2004

Usually, any geographical population is partitioned at several levels, which are often nested; for instance, Spain is divided in 19 Communities, where each Community is divided in several provinces, each province in counties, and each county is divided in administrative districts. Thus a design-based estimator may be accurate enough when calculated at the Community or the province levels, but its accuracy may become unacceptable for counties and districts. In that cases, model-based estimators decrease their mean square error by using auxiliary information in the form of regression models. Here we are interested in observing and analyzing the effect of decreasing the level of aggregation (i.e., decrease the rate *total sample size/number of domains*) in the behaviour of some design-based estimators and some model-based estimators, in order to know till which level are design-based estimators reliable and when is necessary to attend to model-based estimators. Further, it is reasonable to suppose that as long as we decrease area sizes, extra "useful" auxiliary information is needed, so models with more information should provide better estimators of small areas. Models with small area random effects are becoming rather popular. Here we compare the performance of fixed effects models and random effects models for the simulated data.

For this purpose, an artificial population is generated imitating a census data file of some geographical population. This file contains the variables defining domains or areas at six nested levels of aggregation, a variable used for stratification, three auxiliary variables, one of them categorical, and the target variable. From this artificial population, 10,000 samples have been extracted. For each sample, small-area estimators have been calculated at each level of aggregation. At the end, two efficiency measures have been computed for each estimator and each level of aggregation: the first measuring the bias, and the second the mean squared error. The evolution of such measures is investigated when decreasing the level of aggregation.

2 The artificial population

The artificial population is a data file with 11 variables and 300,000 records. Each record represents a household of an imaginary country. The file is generated with the purpose of simulating surveys on income and living conditions. See the description of the file in Table 2.1. The first 7 variables are geographical characteristics, where D_1-D_6 have a nested structure and define the domains or areas, while H, representing strata, produces cross-sections with D_1-D_6. The last 4 variables represent household characteristics. For variable G (socioeconomic condition group), we assume that labour activities are classified into two groups: "better paid" and "worse paid", denoting the first group by BPA.

Table 2.1: Description of the Artificial Population.

variable	position	name and description	values
		geographical characteristics	
D_1	1	*Region*	1–8
D_2	2–3	*Community*	1–16
D_3	4–5	*Province*	1–32
D_4	6–7	*County*	1–64
D_5	8–10	*District*	1–128
D_6	11–13	*Zone*	1–256
H	14	*Stratum*	1–6
		household characteristics	
X_1	15–16	*Total number of household members*	01–30
X_2	17–20	*Total area of the dwelling (m^2)*	0000–9999
G	21	*Socioeconomic condition group*	1–4
		All members of the household are unemployed	1
		There are employed members. None of them in BPA	2
		There are employed members, but only one in BPA	3
		There are employed members. Two or more in BPA	4
		target variable	
Y	22–26	*Total net monetary annual income of the household*	00000–99999

Generation of stratum-zone sizes

Let N_{hd_6} be the number of households on stratum h and zone d_6, $h = 1,\dots,6$, $d_6 = 1,\dots,256$. These numbers are generated according to the following algorithm

1. Generate $6 \times 256 = 1{,}536$ uniform numbers in the interval $(0,1)$. Denote these numbers by u_{hd_6}, $h = 1,\dots,6$, $d_6 = 1,\dots,256$.
2. Calculate $u = \sum_{h=1}^{6}\sum_{d_6=1}^{256} u_{hd_6}$ and $v_{hd_6} = u_{hd_6}/u$, $h = 1,\dots,6$, $d_6 = 1,\dots,256$.
3. Calculate $N_{hd_6} = [300{,}000\, v_{hd_6}]$, $h = 1,\dots,6$, $d_6 = 1,\dots,256$, where $[\cdot]$ denotes "integer part".
4. If $\sum_{h=1}^{6}\sum_{d_6=1}^{256} N_{hd_6} = 300{,}000 - n$, with $n > 0$, then add one to the first n sizes N_{hd_6}. Use lexicographic order in subindexes (h, d_6).
5. If $\sum_{h=1}^{6}\sum_{d_6=1}^{256} N_{hd_6} = 300{,}000$, then stop.

Generation of geographical characteristics

Imputation of numerical values to variables H and D_6 is done by assigning sequentially $H = h$ and $D_6 = d_6$ to N_{hd_6} records, $h = 1,\dots,6$, $d_6 = 1,\dots,256$. Variable D_5 is calculated from D_6 by applying formula

$$D_5 = \left[\frac{D_6 + 1}{2}\right].$$

Similar formulas are used to generate D_4, D_3, D_2 and D_1, i.e.

$$D_4 = \left[\frac{D_5 + 1}{2}\right], \quad D_3 = \left[\frac{D_4 + 1}{2}\right], \quad D_2 = \left[\frac{D_3 + 1}{2}\right], \quad D_1 = \left[\frac{D_2 + 1}{2}\right].$$

In this way, the number of areas are doubled from D_ℓ to $D_{\ell+1}$, which means that sample sizes are approximately divided by two.

Generation of household characteristics

- X_1 is generated by

$$\begin{cases} X_1 \sim Poisson(\lambda_{hd_6}) + 1 & \text{if} \quad Poisson(\lambda_{hd_6}) + 1 < 30 \\ X_1 = 30 & \text{otherwise}, \end{cases}$$

  where $\lambda_{hd_6} = 0.8 + 1.5U + h/6 + d_6/256$, $h = 1, \ldots, 6$, $d_6 = 1, \ldots, 256$ and $U \sim Uniform(0, 1)$.
- X_2 is simulated from

$$X_1 \sim \mathcal{N}(\mu_{hd_6}, \sigma^2_{hd_6}),$$

  where

$$\mu_{hd_6} = 80 + 20U + 2h, \quad \sigma_{hd_6} = 5 + \frac{2d_6}{256}, \quad h = 1, \ldots, 6, \; d_6 = 1, \ldots, 256$$

  and $U \sim Uniform(0, 1)$.
- G is simulated, conditionally on X_1, from the following discrete distributions:

$$G|_{X_1=1} \sim \begin{pmatrix} 0.1 & 0.5 & 0.4 \\ 1 & 2 & 3 \end{pmatrix} \quad \text{and} \quad G|_{X_1=2,3,\ldots} \sim \begin{pmatrix} 0.05 & 0.45 & 0.4 & 0.1 \\ 1 & 2 & 3 & 4 \end{pmatrix}.$$

  This is to say that for the variable G, two cases are considered. If $X_1 = 1$, G is simulated from a discrete distribution taking values 1, 2, 3 with probabilities 0.1, 0.5 and 0.4 respectively. If $X_1 = 2, 3, \ldots$, G is simulated from a discrete distribution taking values 1, 2, 3, 4 with probabilities 0.05, 0.45, 0.4 and 0.1 respectively.

Generation of target variable

Y is simulated from the normal mixed model

$$Y_{hd_6gj} = u_{d_6} + a_h + b_g + \beta_1 X_{hd_6gj1} + \beta_2 X_{hd_6gj2} + e_{hd_6gj}, \qquad (2.1)$$
$$h = 1, \ldots, 6, \; d_6 = 1, \ldots, 256, \; g = 1, \ldots, 4, \; j = 1, \ldots, N_{hd_6g},$$

where u_{d_6} and e_{hd_6gj} are the zone and household level residuals, which are independent random variables with distributions $\mathcal{N}(0, \sigma_u^2)$ and $\mathcal{N}(0, \sigma_e^2)$ respectively. Indexes h, d_6, g and j are used to denote stratum, zone, socioeconomic group and household respectively. Therefore, N_{hd_6g} is the number of households in stratum h, zone d_6 and group g, and Y_{hd_6gj}, X_{hd_6gj1}, X_{hd_6gj2} are the values that Y, X_1, X_2 take on the household j of the group g, zone d_6 and stratum h.

The following parameter values are used to generate the artificial population: $\beta_1 = 500$, $\beta_2 = 25$, $\sigma_u^2 = 1000$, $\sigma_e^2 = 750$, $a_h = 4000 + 300h$, $h = 1, \ldots, 6$, and $b_g = 5000 + 500g$, $g = 1, \ldots, 4$.

Target parameters

Let us use index d_ℓ for geographical characteristic D_ℓ, $\ell = 1, \ldots, 6$. Target parameters are

$$\overline{Y}_{d_\ell} = \frac{1}{N_{d_\ell}} \sum_{h=1}^{6} \sum_{g=1}^{4} \sum_{j=1}^{N_{hd_\ell g}} Y_{hd_\ell gj}, \quad \ell = 1, \ldots, 6,$$

where $N_{hd_\ell g}$ and $Y_{hd_\ell gj}$ are defined in the same way as $N_{hd_6 g}$ and $Y_{hd_6 gj}$, and N_{d_ℓ} is the number of households in domain $D_\ell = d_\ell$.

3 Notation and estimators

3.1 Notation

The following notation is used

- *Indexes:* s is used for sample and r for nonsample, $d = 1, \ldots, D$ for small areas defined by one of the variables $D_1 - D_6$, $g = 1, \ldots, G$ for socioeconomic group, and finally $j = 1, \ldots, n$ for households.
- *Sizes:* N for population and n for sample. When N or n have indexes they denote size of the corresponding indexed set. For example, n_h is the sample size of stratum h.
- *Totals:* Y or X. When Y or X have indexes they denote the total of the corresponding indexed set. For example, Y_d denotes the total in small area d.
- *Means:* $\overline{Y}$ or $\overline{X}$. When $\overline{Y}$ or $\overline{X}$ have indexes they denote the mean of the corresponding indexed set. For example, $\overline{Y}_d$ denotes the mean of small area d.
- *Weights:* w_j is used for household j and is defined as the inverse of the inclusion probability of household j in the sample. Also, when w has indexes it denotes the sum of weights of the corresponding indexed set.

3.2 Design-based estimators

The following design-based estimators (see e.g. Särndal, Swensson and Wretman (1992)) are considered:

- *w-direct estimator:* It is the classical direct estimator.

$$\widehat{\overline{Y}}_d^{wdirect} = \frac{\sum_{j \in s \cap d} w_j Y_j}{\widehat{N}_d^{direct}}, \quad \widehat{N}_d^{direct} = \sum_{j \in s \cap d} w_j .$$

- *Basic synthetic estimator:* It relies on the idea that the population is partitioned in groups larger than areas, for which direct estimators with good precision are available. It is approximatelly unbiased when all areas contained in a group have the same mean as the whole group. This estimator was used by the USA National

Center for Health Statistics in 1968.

$$\widehat{\overline{Y}}_d^{synt} = \frac{1}{N_d}\sum_{g=1}^{G} N_{dg}\widehat{\overline{Y}}_g^{wdirect} = \frac{1}{N_d}\sum_{g=1}^{G} N_{dg}\left(\frac{\sum_{j\in s\cap g} w_j Y_j}{\sum_{j\in s\cap g} w_j}\right).$$

- *Sample size dependent estimator:* It is constructed by composition of the *w-direct* and the *basic synthetic* estimators. It was proposed by Drew, Sigh and Chouldry (1982).

$$\widehat{\overline{Y}}_d^{ssd} = \gamma_d \widehat{\overline{Y}}_d^{wdirect} + (1-\gamma_d)\widehat{\overline{Y}}_d^{synt},$$

where γ_d is calculated by the following formula

$$\gamma_d = \begin{cases} 1 & \text{if } \widehat{N}_d^{direct} \geq N_d, \\ \frac{\widehat{N}_d^{direct}}{N_d} & \text{otherwise.} \end{cases}$$

3.3 Generalized regression estimators

These estimators arise from fitting a model with general shape

$$y = X\boldsymbol{\beta} + W^{-1/2}\mathbf{e}, \tag{3.1}$$

(see e.g. Section 10.5 of Särndal, Swensson and Wretman (1992)) where $\boldsymbol{y}$ is the vector of sample observations of the target variable Y, $\boldsymbol{X}$ is the matrix whose columns are the observations of auxiliary variables, $\boldsymbol{\beta}$ is the vector of coefficients of mentioned variables, W is a diagonal matrix of known positive elements, and $\boldsymbol{e}$ is the vector of individual errors, satisfying $\boldsymbol{e} \sim N(0, \sigma_e^2 I_n)$, where I_n denotes the identity matrix of size n. Fitting this model by weighted least squares, we get individual predictions

$$\widehat{Y}_{dj} = \boldsymbol{x}_{dj}\widehat{\boldsymbol{\beta}}, \; d = 1,\ldots,D, \; j = 1,\ldots,N_d, \tag{3.2}$$

where $\boldsymbol{x}_{dj}$ represents the row of matrix $\boldsymbol{X}$ corresponding to household j of area d, and

$$\widehat{\boldsymbol{\beta}} = (\boldsymbol{X}^t W \boldsymbol{X})^{-1}\boldsymbol{X}^t W \boldsymbol{y}. \tag{3.3}$$

Then, the prediction of area mean

$$\overline{Y}_d = \frac{1}{N_d}\sum_{j=1}^{N_d} Y_{dj}$$

is the mean of predictions of individual values

$$\widehat{\overline{Y}}_d = \frac{1}{N_d}\sum_{j=1}^{N_d} \widehat{Y}_{dj}. \tag{3.4}$$

We have used four specific models of the form (3.1), starting from a simple model, and sequentially including more information to the model. Each model provides a generalized regression estimator.

- *Generalized regression synthetic estimator (gsynt):* It is based on the model with common intercept α and X_1 (number of members of the household) as explanatory variable, i.e.

$$y_{dj} = \alpha + \beta x_{dj1} + w_{dj}^{-1/2} e_{dj}, \; d = 1, \dots, D, \; j = 1, \dots, n_d. \tag{3.5}$$

This model is a particular case of model (3.2) taking $\beta = (\alpha, \beta)^t$ and $\mathbf{x}_{dj} = (1, x_{dj1})$. Applying formula (3.3), estimator $\widehat{\beta} = (\widehat{\alpha}, \widehat{\beta})^t$ is obtained, where $\widehat{\alpha} = \overline{Y}^{\widehat{wdirect}} - \widehat{\beta}\overline{X}_1^{\widehat{wdirect}}$. Replacing $\widehat{\alpha}$ in predictions $\widehat{Y}_{dj} = \widehat{\alpha} + \widehat{\beta} x_{dj1}$, we get

$$\widehat{Y}_{dj} = \overline{Y}^{\widehat{wdirect}} + \widehat{\beta}(x_{dj1} - \overline{X}_1^{\widehat{wdirect}}).$$

Making the average of predictions as in (3.4), we get the following expression of the $gsynt$ estimator

$$\overline{Y}_d^{\widehat{gsynt}} = \overline{Y}^{\widehat{wdirect}} + \widehat{\beta}(\overline{X}_{d1} - \overline{X}_1^{\widehat{wdirect}}).$$

- *Generalized regression estimator 1 (greg1):* Here the model is built by replacing the common intercept of model (3.5) by area fixed effects u_d, i.e.

$$y_{dj} = u_d + \beta x_{dj1} + w_{dj}^{-1/2} e_{dj}, \; d = 1, \dots, D, \; j = 1, \dots, n_d. \tag{3.6}$$

Here $\beta = (u_1, \dots, u_{d-1}, u_d, u_{d+1}, \dots, u_D, \beta)^t$ and $\mathbf{x}_{dj} = (0, \dots, 0, 1, 0, \dots, 0, x_{dj1})$. Again, using formula (3.3), estimators of u_d, $d = 1, \dots, D$ and β are obtained. Replacing formulas of estimators $\widehat{u}_d$, $d = 1, \dots, D$ in individual predictions (3.2), and averaging, we get the following expression of the $greg1$ estimator

$$\overline{Y}_d^{\widehat{greg1}} = \overline{Y}_d^{\widehat{wdirect}} + \widehat{\beta}(\overline{X}_{d1} - \overline{X}_{d1}^{\widehat{wdirect}}).$$

- *Generalized regression estimator 2 (greg2):* In this case the model is obtained from (3.6) by incorporating a second explicative variable, X_2 (total area of the dwelling in m^2), so that

$$y_{dj} = u_d + \beta_1 x_{dj1} + \beta_2 x_{dj2} + w_{dj}^{-1/2} e_{dj}, \; d = 1, \dots, D, \; j = 1, \dots, n_d. \tag{3.7}$$

By the same procedure as before, we get the $greg2$ estimator

$$\overline{Y}_d^{\widehat{greg2}} = \overline{Y}_d^{\widehat{wdirect}} + \widehat{\beta}_1(\overline{X}_{d1} - \overline{X}_{d1}^{\widehat{wdirect}}) + \widehat{\beta}_2(\overline{X}_{d2} - \overline{X}_{d2}^{\widehat{wdirect}}),$$

- *Generalized regression estimator 3 (greg3):* This estimator is based on the model

$$y_{dgj} = u_d + b_g + \beta_1 x_{dgj1} + \beta_2 x_{dgj2} + w_{dgj}^{-1/2} e_{dgj}, \tag{3.8}$$

where b_g is the effect of socioeconomic group g, $g = 1, \dots, G-1$ ($b_G = 0, G = 4$). The estimator of the dth small area mean can be expressed as

$$\overline{Y}_d^{\widehat{greg3}} = \overline{Y}_d^{\widehat{wdirect}} + \sum_{g=1}^{G-1} \widehat{b}_g \left(\frac{N_{dg}}{N_d} - \frac{w_{dg}}{w_d} \right) + \widehat{\beta}_1(\overline{X}_{d1} - \overline{X}_{d1}^{\widehat{wdirect}}) + \widehat{\beta}_2(\overline{X}_{d2} - \overline{X}_{d2}^{\widehat{wdirect}}),$$

where $\widehat{b}_g$, $\widehat{\beta}_1$ and $\widehat{\beta}_2$ are the weighted least squares estimators obtained by (3.3).

3.4 Empirical best linear unbiased estimators

We consider estimators obtained by fitting to the sample a random effects model of the form

$$y = X\beta + Zu + W^{-1/2}e\,, \tag{3.9}$$

where $y = y_{n\times 1}$, $X = X_{n\times p}$, $\beta = \beta_{p\times 1}$, $Z = Z_{n\times D} = \text{diag}(\mathbf{1}_{n_1},\ldots,\mathbf{1}_{n_D})$ with $\mathbf{1}_a^t = (1,\ldots,1)_{1\times a}$ and $W = W_{n\times n} = \text{diag}(w_{11},\ldots,w_{Dn_D})$ with $w_{11} > 0,\ldots,w_{Dn_D} > 0$ known. Vectors of area and household random effects, $u = u_{D\times 1} \sim N(0,\sigma_u^2 I_D)$ and $e = e_{n\times 1} \sim N(0,\sigma_e^2 I_n)$ respectively, are assumed to be independent with unknown variance components σ_u^2 and σ_e^2. Under this model, the variance-covariance matrix of y is

$$V = \sigma_u^2 ZZ^t + \sigma_e^2 W^{-1}\,.$$

The individual predictions of non sampled units are

$$\widehat{Y}_{dj} = x_{dj}\widehat{\beta} + \gamma_d^w\left(\widehat{\overline{Y}}_d^{\,wdirect} - \widehat{\overline{X}}_d^{\,wdirect}\widehat{\beta}\right),$$

where

$$\widehat{\beta} = (X^tV^{-1}X)^{-1}X^tV^{-1}y,\quad \gamma_d^w = \frac{\sigma_u^2}{\sigma_u^2+\sigma_e^2/w_d},\quad w_d = \sum_{j=1}^{n_d} w_{dj},\quad d=1,\ldots,D.$$

The *blup* of small area mean $\overline{Y}_d$ is obtained under the assumption of known variance components (see Chapter 2 of Vaillant et al (2000) for an overview of the Prediction Theory) by

$$\begin{aligned}\widehat{\overline{Y}}_d &= \frac{1}{N_d}\left(\sum_{j\in s\cap d} Y_{dj} + \sum_{j\in r\cap d}\widehat{Y}_{dj}\right)\\ &= (1-f_d)\left[\overline{X}_d\widehat{\beta} + \gamma_d^w\left(\widehat{\overline{Y}}_d^{\,wdirect} - \widehat{\overline{X}}_d^{\,wdirect}\widehat{\beta}\right)\right] + f_d\left[\widehat{\overline{y}}_d + (\overline{X}_d - \widehat{\overline{X}}_d)\widehat{\beta}\right],\end{aligned} \tag{3.10}$$

where $\overline{X}_d = \frac{1}{N_d}\sum_{j=1}^{N_d} x_{dj}$, $\widehat{\overline{X}}_d = \frac{1}{n_d}\sum_{j=1}^{n_d} x_{dj}$ and $\widehat{\overline{y}}_d = \frac{1}{n_d}\sum_{j=1}^{n_d} y_{dj}$.

Observe that estimator $\widehat{\overline{Y}}_d$ depends on the variance components through γ_d^w and V^{-1}. By plugging in (3.10) suitable estimators of the variance components, the *empirical blup* (*eblup*) of small area mean $\overline{Y}_d$ is obtained (see Battesse et al. (1988) and Prasad and Rao (1990)). We present some estimators obtained from specific models of the type (3.9).

- *ebluph1:* It uses X_1 as auxiliary variable, so that $x_{dj} = x_{dj1}$ and $\beta = \beta$. Estimated variance components $\widehat{\sigma}_u^2$ and $\widehat{\sigma}_e^2$ are obtained by Henderson's method 3 (see Henderson (1953) or Searle et al (1992)). This small area estimator has been studied by Rao and Choudhry (1995).
- *eblup1:* The only difference of this estimator with respect to *ebluph1* is that β, σ_u^2 and σ_e^2 are estimated by maximizing the likelihood of model (3.9). These

maximum likelihood estimates (MLE) are calculated numerically by using the Fisher-scoring algorithm. We will check which of both variance components estimation methods (maximum likelihood or Henderson method 3) provide better small area estimator.

- *eblup2:* It uses X_1 and X_2 as auxiliary variables. Estimators of model parameters, $\widehat{\beta}_1, \widehat{\beta}_2, \widehat{\sigma}_u^2, \widehat{\sigma}_e^2$, are MLE's and they are calculated via Fisher-scoring algorithm.
- *eblup3:* It uses X_1, X_2 and G as auxiliary variables. Estimators of model parameters, $\widehat{b}_g, \widehat{\beta}_1, \widehat{\beta}_2, \widehat{\sigma}_u^2, \widehat{\sigma}_e^2$ are MLE's and they are computed by using Fisher-scoring algorithm.

4 Measures of sampling errors in simulation experiment

In order to evaluate the precision and accuracy of proposed small area estimators for estimating the average net income, $\overline{Y}_d$, K samples are drawn from the artificial population and estimations are obtained for each sample. Let $\widehat{\overline{Y}}_d(k)$ be the estimate of the mean $\overline{Y}_d$ for the small area d in the k-th replicated sample. The following standard performance criteria are considered:

1. The *average relative bias* for small area d

$$ARB_d = \frac{1}{K}\sum_{k=1}^{K}\left(\frac{\widehat{\overline{Y}}_d(k)}{\overline{Y}_d}\right)100. \tag{4.1}$$

2. The *relative mean squared error* for small area d

$$RMSE_d = \frac{100}{\overline{Y}_d}\sqrt{\frac{1}{K}\sum_{k=1}^{K}\left(\widehat{\overline{Y}}_d(k) - \overline{Y}_d\right)^2}. \tag{4.2}$$

5 Monte Carlo simulation experiment

A C++ Builder program has been developed to extract random samples from the data file and to evaluate estimators and performance measures. The number of replications of the simulation experiment is $K = 10{,}000$. A *stratified sampling design*, with simple random sampling without replacement inside each of the strata and total sample size $n = 600$, has been used. Population sizes of strata N_h are calculated from the artificial population, and sampling weights w_h have been taken from the Spanish Family Budget Survey for a province with average size. From these two quantities, by the relation $w_h = N_h/n_h$, sample sizes inside each stratum n_h have been derived. These quantities are shown in Table 5.1.

***Table** 5.1: Sizes and weights per stratum.*

stratum	1	2	3	4	5	6
N_h	49828	52717	48051	48865	48831	51708
n_h	100	105	96	98	98	103
w_h	498.28	502.07	500.53	498.62	498.28	502.02

With the obtained sample, all estimators of the average net income $\overline{Y}_d$ are calculated for small areas defined by $D_1, \ldots, D_6$. When the process of $K = 10,000$ replications is finished, efficiency measures are evaluated.

In order to clarify the role that sample size ($n = 600$) and number of small areas (D) play in the analysis of the numerical results, in Table 5.2 we present the quantities n/D for D_1–D_6.

***Table** 5.2: Average sample sizes per small areas.*

	D_1	D_2	D_3	D_4	D_5	D_6
D	8	16	32	64	128	256
$600/D$	75	37.5	18.75	9.375	4.6875	2.34375

***Table** 5.3: Means over small areas and standard deviations (in brackets) of ARB_d.*

Estimator	D_1	D_2	D_3	D_4	D_5	D_6
wdirect	99.998	100.005	100.026	100.001	100.001	100.145
	(0.009)	(0.022)	(3.167)	(0.036)	(0.107)	(0.966)
synt	100.012	100.037	100.027	100.143	100.337	100.704
	(2.012)	(2.595)	(3.127)	(4.207)	(5.606)	(7.318)
ssd	100.012	100.037	100.026	100.020	100.062	100.219
	(2.068)	(2.640)	(3.163)	(0.551)	(1.036)	(1.968)
rsynt	100.015	100.036	100.038	100.161	100.343	100.693
	(1.371)	(2.160)	(2.758)	(3.972)	(5.404)	(7.120)
greg1	100.002	99.999	100.002	100.005	100.004	100.142
	(0.008)	(0.018)	(0.025)	(0.031)	(0.107)	(0.915)
greg2	100.002	99.999	100.003	100.006	100.006	100.142
	(0.005)	(0.016)	(0.023)	(0.029)	(0.108)	(0.895)
greg3	100.002	99.999	100.002	100.004	100.001	100.145
	(0.006)	(0.017)	(0.024)	(0.030)	(0.105)	(0.966)
ebluph1	100.000	99.994	99.992	99.972	98.890	90.194
	(0.008)	(0.018)	(0.025)	(0.038)	(0.918)	(6.423)
eblup1	99.984	99.975	99.938	99.868	98.740	90.171
	(0.010)	(0.013)	(0.023)	(0.044)	(0.889)	(6.382)
eblup2	99.977	99.960	99.885	99.601	99.680	99.827
	(0.008)	(0.013)	(0.030)	(1.529)	(2.232)	(3.035)
eblup3	100.001	100.026	100.020	100.046	100.113	100.264
	(0.595)	(0.767)	(1.028)	(1.331)	(1.760)	(2.336)

In Table 5.3, means and standard deviations (in brackets) of ARB_d over small areas, that is, $\overline{ARB} = D^{-1} \sum_{d=1}^{D} ARB_d$ and $S_{ARB} = \left[D^{-1} \sum_{d=1}^{D}(ARB_d - \overline{ARB})^2\right]^{1/2}$, are listed for each variable defining small areas $D_1 - D_6$. In Table 5.4 means and standard deviations of $RMSE_d$ are given. Observe that in Table 5.3, standard deviations provide more information about the amount of bias than the mean, which is "average" bias.

***Table** 5.4: Means over small areas and standard deviations (in brackets) of $RMSE_d$.*

Estimator	D_1	D_2	D_3	D_4	D_5	D_6
wdirect	1.320	1.873	2.697	3.805	5.420	6.687
	(0.083)	(0.152)	(1.647)	(0.488)	(0.917)	(1.000)
synt	1.745	2.101	2.659	3.324	4.572	6.065
	(0.875)	(1.486)	(1.631)	(2.582)	(3.263)	(4.159)
ssd	1.801	2.148	2.693	3.267	4.319	5.575
	(0.884)	(1.497)	(1.646)	(0.386)	(0.726)	(1.275)
rsynt	1.187	1.707	2.303	3.172	4.465	5.823
	(0.655)	(1.306)	(1.508)	(2.392)	(3.061)	(4.157)
greg1	1.107	1.558	2.211	3.074	4.246	5.087
	(0.089)	(0.160)	(0.295)	(0.502)	(0.927)	(0.925)
greg2	1.068	1.495	2.118	2.931	4.008	4.756
	(0.091)	(0.162)	(0.314)	(0.524)	(0.938)	(0.890)
greg3	1.033	1.446	2.048	2.827	5.132	6.687
	(0.092)	(0.165)	(0.320)	(0.535)	(0.900)	(1.000)
ebluph1	1.107	1.558	2.211	3.218	9.868	28.056
	(0.008)	(0.018)	(0.025)	(0.038)	(0.918)	(6.423)
eblup1	1.110	1.561	2.205	3.214	9.860	27.697
	(0.092)	(0.158)	(0.287)	(0.575)	(3.362)	(8.234)
eblup2	1.067	1.503	2.132	2.537	3.406	4.448
	(0.093)	(0.162)	(0.303)	(0.722)	(0.998)	(1.421)
eblup3	0.913	1.240	1.642	2.220	2.976	3.842
	(0.241)	(0.314)	(0.380)	(0.582)	(0.851)	(1.263)

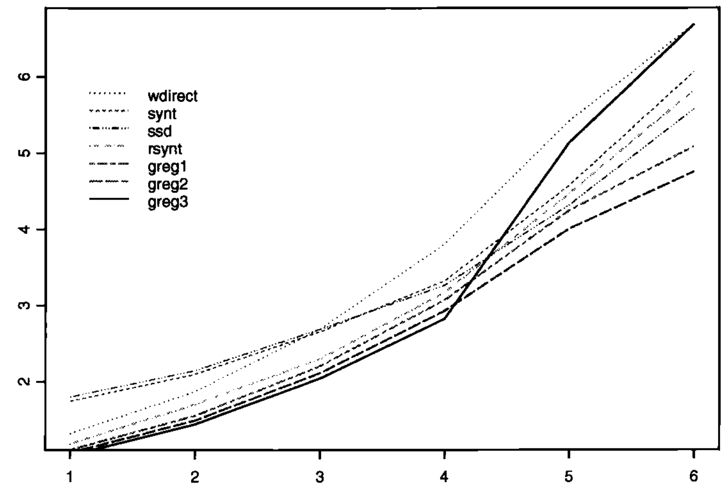


***Figure 1**: $\overline{RMSE}$ of design-based and greg estimators, for D_1-D_6.*

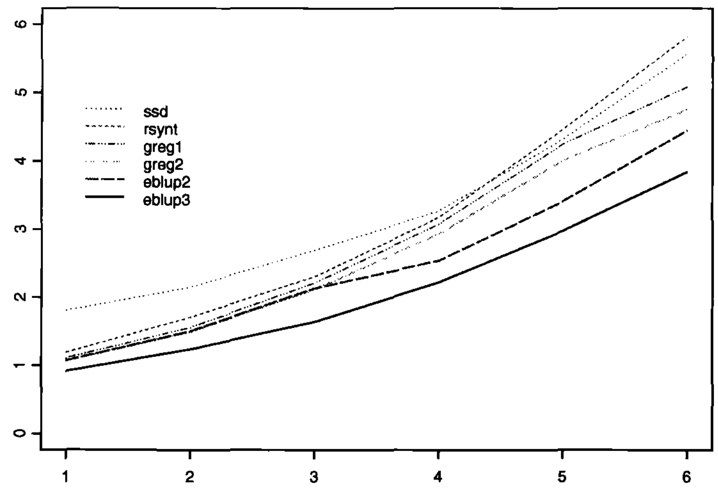


***Figure 2**: $\overline{RMSE}$ of best estimators, for D_1-D_6.*

The most interesting estimators are the ones with better performance for $D_4 - D_6$, whose average area sample sizes are smaller than 10 units. We can observe better the evolution of $\overline{RMSE}$ for the small areas defined by D_1-D_6 in Figures 5.1 and 5.2. Figure 5.1 shows the $\overline{RMSE}$ for design-based and *greg* estimators, while Figure 5.2 represents the same efficiency measure only for the group of estimators with a reasonable behaviour. We make a comparison of estimators one by one.

The w-direct estimator (*wdirect*) uses just data of the target variable in the given small area. As expected, it behaves well with respect to bias in all cases; see that the standard deviation remains small even for D_6. However, its behaviour is not very good in relative mean squared error, since it increases monotonically, differing considerably from the group of good estimators from D_2 on. Thus, the use of *wdirect* can only be recommended for sample sizes larger than 40.

The basic synthetic estimator (*synt*), which uses as auxiliary information just the socioeconomic group G, is relatively stable in $\overline{RMSE}$ with respect to n/D. But we can see in Table 5.3 that it is quite biased in all cases, being the standard deviation of the ARB_d unacceptably large for D_4–D_6.

The sample size dependent estimator (*ssd*) is a composition of the direct and the basic synthetic estimator. Surprisingly, this estimator gets quite good results in both performance measures, relative mean squared error and average relative bias, being its $\overline{RMSE}$ comparably to the *greg1* and *greg2* estimators, and being its bias smaller that all *eblup* estimators, as expected, and also better than both synthetic and generalized regression synthetic estimator. This estimator is a good alternative, either when no more than a grouping variable is available as auxiliary information, or when it is not found a good model fitting the data.

The generalized regression synthetic estimator (*rsynt*) clearly improves the numerical results of *wdirect* and *synt* estimators, improving the direct estimator even for larger areas. The reason is that model (3.5) fits reasonably well to data. However, it has comparable amount of bias to the basic synthetic estimator (see the standard deviation).

Comparing generalized regression estimators, we can see in *greg1* and *greg2* how extra information provide a decrease in $\overline{ARB}$ and $\overline{RMSE}$, except when we arrive to *greg3*, whose $\overline{RMSE}$ increases quickly when going from D_4 on. We can look at case D_6 to explain this phenomenon. For small areas defined by D_6, model (3.8) from *greg3* has $D + (G - 1) + 2 = 261$ regression parameters. Since sample size is 600, there are $600/261 \approx 2.3$ observations per parameter. This ratio is too small in order to estimate model parameters with low standard deviations. Under these conditions, these model is not stable and therefore its use is not recommended.

Now we compare model-based estimators. Looking at Table 5.4, we see that *ebluph1* and *eblup1* increase considerably their $\overline{RMSE}$ for D_5 and D_6, being *greg1* (based in the same model but with fixed effects) much preferable. Since sampling fractions n_d/N_d are close to zero, the eblup estimators are approximately given by

$$\widehat{\overline{Y}}_d \cong \overline{X}_d\widehat{\beta} + \gamma_d^w \left(\widehat{\overline{Y}}_d^{wdirect} - \widehat{\overline{X}}_d^{wdirect}\widehat{\beta}\right). \tag{5.3}$$

The problem here is that the variability of X_1 in the small number of observations is not sufficient to estimate with precision the variability between households σ_e^2 and between areas σ_u^2. This fact causes a strong negative bias on the synthetic part of the *eblup1* estimator, $\overline{X}_d\widehat{\beta}$. In fact, for D_6, its $\overline{ARB}$ is 85.897, and its standard deviation is 1.587. This synthetic estimator is corrected by the second term on the right of (5.3), but this term is affected by the same problem, making the correction rather poor. In this experiment, estimators based on linear models with small area random effects appear to be very sensible to goodness–of–fit, providing worse results than estimators based on linear models with small area fixed effects, but also worse than design-based estimators which do not make use of any covariate. It is interesting to note that Fisher–scoring algorithm (*eblup1*) is slightly better that Henderson's method 3 (*ebluph1*).

Estimators *greg2* and *eblup2* rely on the same linear regression models. The difference is that small area effects are fixed in the first case and random in the second case. We can see that for sample sizes smaller than 20, *eblup2* has less $\overline{RMSE}$. This indicates that as soon as selected model fits better to data, their corresponding estimators perform more efficiently.

The same occurs with *greg3* and *eblup3*. Since the model of *eblup3* is very close to the real one, this estimator presents the best numerical results in the simulation experiment. Mean squared errors remain small even in the case D_6.

6 Summary and Conclusions

In practical applications of the estimation of means and totals for small areas, statisticians needs recommendations about what type of estimator to use, when it is better a model-based approach than a model-assisted one, or how estimators are affected by the ratio *sample size/number of domains*. Theoretical properties of estimators give answers to these questions under ideal conditions, i.e. if sufficient hypotheses are fulfilled and/or sample sizes are large enough. However, in practice, models do not perfectly fit to data and sample sizes are small. Then simulation studies play an important role to gain intuition and obtain conclusions about the behaviour of estimators. The simulation study presented in this paper has tried to reproduce artificially populations and sampling designs appearing in official statistics, so that our conclusions are valuable for applied statisticians and can be taken into account by Statistical Offices.

From the developed simulation study, under the artificial population generated as described in Section 2, we can extract the following conclusions:

1. If there is an informative grouping variable available, a good choice for estimating means of areas with average sample sizes smaller than 20 is the *ssd* estimator.
2. If there is available at least one "good" covariable, its use is always recommended.
3. Best numerical results are obtained for those estimators with models fitting "well" to data. A bad model produces a bad estimator.
4. When a good model is not found, it is better to use models with fixed effects. Among estimators based in models with random effects, the worst numerical results are obtained by *ebluph1* and *eblup1*. These two estimators behave acceptably well only for average sample sizes greater than 10.
5. If a good model is available, then random effects are clearly preferred to fixed effects for sample sizes smaller than 20. Note that there is a significant decrease of relative mean squared error when passing from *greg2* to *eblup2* or from *greg3* to *eblup3*.

Acknowledgements

This work has been done in the set up of a R&D contract with Instituto Nacional de Estadística (INE). It has been also partially supported by grant BMF2003-04820 of the Spanish Ministery of Science and Technology. Ideas appearing in the manuscript come from the collaboration of authors with INE and with EURAREA team. EURAREA (Enhancing Small Area Estimation Techniques to Meet European Needs, IST-2000-5.1.8, 2001-2003) is a research project funded by EUROSTAT under the Fifth Frame work Programme.

References

Battesse, G. E., Harter, R. M. and Fuller, W. A. (1988). An error-component model for prediction of county crop areas using survey and satellite data. *Journal of the American Statistical Association*, 83, 28-36.

Drew, J. D., Singh, M. P. and Choudhry, G. H. (1982). Evaluation of small area techniques for the Canadian Labour Force Survey. *Survey Methodology*, 8, 17-47.

Ghosh, M. and Rao, J. N. K. (1994). Small area estimation: an appraisal. *Statistical Science*, 9, 55-93.

Henderson, C. R. (1953). Estimation of variance and covariance components. *Biometrics*, 9, 226-252.

Prasad, N. G. N. and Rao, J. N. K. (1990). The estimation of the mean squared error of small-area estimators. *Journal of the American Statistical Association*, 85, 163-171.

Rao, J. N. K. and Choudhry, G. H. (1995). Small area estimation: overview and empirical study. *Business Survey Methods* (Cox, Binder, Chinnappa, Christianson, Colledge, Kott, eds.). John Wiley, 527-542.

Särndal, C., Swensson, B. and Wretman, J. (1992). *Model Assisted Survey Sampling*. Springer-Verlag.

Searle, S. R., Casella, G. and McCullogh, C. E. (1992). *Variance Components*. John Wiley. New York.

Valliant, R., Dorfman, A. H. and Royall, R. M. (2000). *Finite Population Sampling and Inference. A Prediction Approach*. John Wiley. New York.

Book reviews

MOST HONOURABLE REMEMBRANCE.
THE LIFE AND WORK OF THOMAS BAYES

Andrew I. Dale

Springer-Verlag, New York, 2003
668 pages with 29 illustrations

The author of this book is professor of the Department of Mathematical Statistics at the University of Natal in South Africa. Andrew I. Dale is a world known expert in History of Mathematics, and he is also the author of the book "A History of Inverse Probability: From Thomas Bayes to Karl Pearson" (Springer-Verlag, 2^{nd}. Ed., 1999).

The book is very erudite and reflects the wide experience and knowledge of the author not only concerning the history of science but the social history of Europe during XVIII century. The book is appropriate for statisticians and mathematicians, and also for students with interest in the history of science. Chapters 4 to 8 contain the texts of the main works of Thomas Bayes. Each of these chapters has an introduction, the corresponding tract and commentaries. The main works of Thomas Bayes are the following:

Chapter 4: **Divine Benevolence,** *or an attempt to prove that the principal end of the Divine Providence and Government is the happiness of his creatures. Being an answer to a pamphlet, entitled, "Divine Rectitude; or, An Inquiry concerning the Moral Perfections of the Deity". With a refutation of the notions therein advanced concerning Beauty and Order, the reason of punishment, and the necessity of a state of trial antecedent to perfect Happiness.*

This is a work of a theological kind published in 1731. The approaches developed in it are not far away from the rationalist thinking. This is not surprising because Thomas Bayes was a Presbyterian minister in Tunbridge Wells, son of the Rev. Joshua Bayes, a Nonconformist minister. As A.I. Dale comments, "...dissenters (or Nonconformists) grew disenchanted with the Established Church of England. The rise of Natural Philosophy led to an increased interest in Natural (as well as Revealed) Religion, and it was believed that the character of God could (and would) be revealed by a scientific study of His works".

Chapter 5: **An Introduction to the Doctrine of Fluxions**, *and defence of the mathematicians against the objections of the author of the "Analyst", so far as they are designed to affect their general methods of reasoning.*

Published in 1736, in this work Thomas Bayes defended the logical foundation of Newton's calculus against the opinion of George Berkeley, the author of The Analyst. Dale suggests that Bayes was elected as a Fellow of the Royal Society in 1742 on the strength of this work.

Chapter 6: **A Letter from the late Reverend Mr. Thomas Bayes, F.R.S. to John Canton.** It was published in 1763 in *Philosophical Transations of the Royal Society of London*. The fact that Bayes died in 1761 suggests that it was communicated to Canton, only after Bayes's death. This letter is about the behaviour of the Stirling-de Moivre asymptotic series for $\log(z!)$.

Chapter 7: **An Essay towards solving a Problem in the Doctrine of Chances.**

It was published posthumously by his friend Richard Price in the Philosophical Transactions of the Royal Society of London in 1763. Quoting Dale, "In the introduction to the first volume of their *Breakthroughs in Statistics* Kotz and Johnson listed eleven works, up to and including Galton's Natural Inheritance, that have had lasting and fundamental effects on the direction of statistical thought and practice. One is these is Thomas Bayes's *Essay towards solving...*". In this Essay he was the first person to formally define a method of calculating the probability of an event occurring based upon the frequency with which that event has occurred in the past.

Chapter 8: **A Demonstration of the Second Rule in the Essay towards the Solution of a Problem in the Doctrine of Chances, published in the Philosophical Transactions, Vol. LIII.** In 1764 Richard Price sent to publish this mathematical appendix or supplement to Bayes's *Essay towards the Solution of a Problem in the Doctrine of Chances*. The first part is apparently due to Bayes himself, but the rest of the results are due to Richard Price.

Chapter 10 comprises several manuscripts of Thomas Bayes of the Library of the Royal Society: 1- a letter to John Canton on infinite series, 2- a letter to John Canton commenting on some remarks by Thomas Simpson on errors in observations and 3- some notes on electricity. It also contains a description of the main papers of Thomas Bayes in the Stanhope collection.

The contents of Chapter 11 are very interesting. They are about an anonymous notebook found in the monument room of the Equitable Life Assurance Society in London. On its first page, dated in 1947 and signed by M.E. Ogborn, it bears the following words: "This book appears to be a mathematical notebook by Rev. Thomas Bayes, F.R.S. The handwriting agrees very well with papers by him in the Canton papers of the Royal

Society, Vol. 2 p. 32." The notebook was written in both longhand and shorthand, and in English, French and Latin. The topics discussed in the notebook are divided by Dale into several groups: 1- mathematics, 2- natural philosophy, 3- celestial mechanics and 4- miscellaneous matters. For instance, one of these "miscellaneous matters" is about the pyramid of Cheops measured by Greaves.

Chapters 2 and 3 are dedicated to the genealogy and life of Thomas Bayes. In a preceding biography, Pearson (1978) commented, "... it is impossible to understand a man's work unless you understand something of his character and unless you understand something of his environment. And his environment means the state of affairs social and political of his own age." Following these premises, Dale has done an excellent job along this book. Thomas Bayes was probably born in 1701 so that in the year 2001 the 300th anniversary of his birth was celebrated. In 1731 he became the Presbyterian Minister at the Meeting House, Mount Sion, Tunbridge Wells in Kent. Many details about his life can be also found in the paper of Bellhouse (2001), "The Reverend Thomas Bayes, FRS: A Biography to Celebrate the Tercentenary of His Birth", *Statistical Science* 2004, Vol. 19, No. 1, 3–43. Little is known about Bayes and he is considered an enigmatic figure of the science. This book is a great contribution to the understanding of this important figure.

Chapter 12, the last, is dedicated to the burial and last wills of Thomas Bayes, and also to describe his tomb that is placed in Bunhill Fields in London, near the Royal Statistical Society head office. Dale gives a description of the inscriptions on the vault of the tomb:

Rev. Thomas Bayes Son of the said Joshua and Ann Bayes (59)
7 April 1761
In recognition of Thomas BayesLs important work in probability
this vault was restored in 1969 with contributions received
from statisticians throughout the world

Pere Puig
Departament de Matemàtiques
Universitat Autònoma de Barcelona

MONTE CARLO STATISTICAL METHODS

Robert CP, Casella G.

2^{nd} ed. Springer: New York, 2004
ISBN 0-387-21239-6
pp. 645 + XXX, 132 illustrations, hardcover
price: 89.95 €

This is the revised second edition of a textbook on statistical methods based on simulation, particularly those based on Markov Chains. Being a textbook, the authors have made an effort to compile, summarize and organize the prolific literature on the field that now has reached a status mature enough to be used as standard techniques for statisticians. The text pays special attention to introduce solidly the main concepts and allows the reader to access easily the more recent developments in the subject. The intended reader should have previous knowledge of statistical inference, but no previous knowledge about Monte Carlo techniques nor Markov Chains theory are needed. Postgraduate students, researchers or practising statisticians aiming to use these techniques will greatly benefit from reading this text. The concepts are clearly explained, with detail when needed, and the theorems have proofs. Also many practical examples are used throughout the book to illustrate the methods. Great effort has been made to enumerate and detail the algorithms. These are printed inside grey-colour boxes for easy identification. Algorithms are explained with pseudo-code to make easy translation to any programming software. No real implementations of the code are done in the book, though the authors mention that the examples were actually implemented in C and references to the BUGS software are given. As a text book intended for classroom use, many problems are proposed at the end of each chapter, without solution. Also short notes covering complementary material are given after the problems section of each chapter. The book has a big deal of references and, though most are prior to the first edition of 1999, many new ones covering recent developments have been added in this second edition.

The book is organized in 14 chapters. The first is a short introduction on likelihood and Bayesian statistical methods, giving motivating examples of difficult problems of inference that can be solved easier by simulation methods. The second chapter introduces the reader into the field of simulation reviewing different methods of random variable generation. For each method, the theory is detailed and algorithms to facilitate the implementation are presented. Two chapters covering the Monte Carlo integration follow this introductory part (introductory but long and rigorous, like the rest of the book). One presents the Monte Carlo methods to approximate univariate and multidimensional integrals. Next discusses methods to estimate and control the variance of the Monte Carlo estimator. The use of Monte Carlo techniques in order to solve optimization problems are presented in chapter 5. Here the authors contrast stochastic methods for exploration (simulation methods to find the maximum) and methods for approximation of the objective function. In this sense, several methods are reviewed like the *simulated annealing* or the *EM algorithm*. Chapter 6 covers the theory of Markov Chains. First we find an explanation of main the results that are needed to establish the convergence of the MCMC techniques. The goal of this initial section is to advise readers who are more interested in the implementations topics of MCMC than in theoretical details. This is followed by a more comprehensive exposition of the theory, including the proofs of the theorems of convergence. The book then progresses through 5 chapters detailing specific techniques of MCMC, with emphasis on algorithms for easy implementation: the general *Metropolis-Hastings algorithm*, the *slice sampler* and the *Gibbs sampler*. For models contemplating a variable number of dimensions, the *reversible jump* algorithms are explained in detail. Chapter 12 presents a selection of useful methods to assess convergence and the notes comment the software CODA for this purpose. Finally, the last two chapters are dedicated to new developments in the field of MCMC, methods that emerge strongly like the *perfect sampling* and *the iterated and sequential importance sampling*.

In summary, this book is attractive to read because it is full of motivating examples. It will be very useful to those interested in implementing the algorithms, because these are clearly explained. The theoretical bases for the methods are covered in detail and this introduces the student or researcher to the background needed to follow the specialized literature of the field, which is prolific.

Víctor Moreno and Raquel Iniesta
Catalan Institute of Oncology

Information for authors and subscribers

Information for authors and subscribers

Submitting articles to SORT

Guidelines for submitting articles

SORT accepts for publication only original articles that have not been submitted to any other journal in the areas of Statistics, Operations Research, Official Statistics and Biometrics. Articles may be either theoretical or applied, and may include computational or educational elements. Publication will be exclusively in English, with an abstract in Catalan (abstract translation into Catalan can be done by the journal's staff).

All articles submitted to thematic sections will be forwarded for systematic peer review by independent specialists and/or members of the Editorial Board, except for those articles specifically invited by the journal or reprinted with permission. Reviewer comments will be sent to the primary author if changes are requested in form or content.

When an original article is received, the journal sends the author a dated receipt; this date will appear in the publication as "reception date". The date in which its final version is received will be the "acceptance date".

To submit an article, the author must send it in PDF format to the electronic address of the journal (sort@idescat.es) especifying his e-mail address and/or a phone number. Once the article has been approved, the author must send a final electronic version, following the instructions of the editor responsible for that article. It is recommended that this final version be submitted using a $\LaTeX 2_\varepsilon$.

In any case, upon request the journal secretary will provide authors with $\LaTeX 2_\varepsilon$ templates and appropriate references to the MSC2000 classification of the American Mathematical Society. For more information about the criteria for using this classification system, authors may consult directly the Mathematics Subject Classification 2000 (MSC2000) at the website of the Institut d'Estadística de Catalunya (http://www.idescat.es/sort/Normes.stm).

Publishing rights and authors' opinions

The publishing rights for SORT belong to the Institut d'Estadística de Catalunya since 1992 through express concession by the Technical University of Catalonia. The journal's current legal registry can be found under the reference number DL B-46.085-1977 and its ISSN is 1696-2281. Partial or total reproduction of this publication is expressly forbidden, as is any manual, mechanical, electronic or other type of storage or transmission without prior written authorisation from the Institut d'Estadística de Catalunya.

Authored articles represent the opinions of the authors; the journal does not necessarily agree with the opinions expressed in authored articles.

Formal guidelines for articles

The first page of the article must contain the following items: title, name of the author(s), professional affiliation and complete address, and an abstract (100-200 words) followed by MSC2000 classification of the American Mathematical Society and the keywords. Before submitting an article, the author(s) would be well advised to ensure that the text uses correct English.

Bibliographic references have to be listed alphabetically at the end of the article, according to the following examples:

Citations
Mahalanobis (1936), Rao (1982b)

Journal articles
Pfeffermann, D. and Barnard, C. H. (1991). Some new estimators for small-area means with application to the assessment of farmland values. *Journal of Business & Economic Statistics*, 9, 73-84.

Books
Platek, R., Rao, J. N. K., Särndal, C. E. and Singh, M. P. (eds.) (1987). *Small Area Statistics: an International Symposium.* New York: John Wiley and Sons.

Parts of books
Lee, E.W., Wei, L.J. and Amato, D.A. (1992). Cox-type regression analysis for large numbers of small groups of correlated failure time observations. In *Survival Analysis: State of the Art*, J.P. Klein and P.K. Goel (eds.), Kluwer Academic Publishers, 237-247.

Web files or "pages"
Nielsen, S. F. (2001). *Proper and improper multiple imputation*
http://www.stat.ku.dk/~feodor/publications/ (10th May 2003).

Multiple publications by a single author are to be listed chronologically. Explanatory notes should be numbered sequentially and placed at the bottom of the corresponding page. Tables and figures should also be numbered sequentially and will be reproduced directly from the submitted originals if it is impossible to include them in the electronic text.

SORT ***(Statistics and Operations Research Transactions)***
Institut d'Estadística de Catalunya (Idescat)
Via Laietana, 58
08003 Barcelona (Spain)
Tel.: +34-93 412 09 24 or +34-93 412 00 88
Fax: +34-93 412 31 45
E-mail: sort@idescat.es

Subscription form

SORT *(Statistics and Operations Research Transactions)*

Name ____
____
Organisation ____
____
Street Address _______________________________________
____
Zip/Postal code ____________________ City ______________
State/Country ______________________ Tel. ______________
Fax ______________ NIF/VAT Registration Number __________
E-mail ____
Date ______________________

Signature

I wish to subscribe to **SORT *(Statistics and Operations Research Transactions)*** for the year 2005 (volume 29)

Annual subscription rates:

– Spain: €22 (VAT included)
– Other countries: €25 (VAT included)

Price for individual issues (current and back issues):

– Spain: €9/issue (VAT included)
– Other countries: €11/issue (VAT included)

Method of payment:

☐ Bank transfer to account number 2013-0100-53-0200698577

☐ Automatic bank withdrawal from the following account number

☐☐☐☐ ☐☐☐☐ ☐☐ ☐☐☐☐☐☐☐☐☐☐

☐ Check made payable to the Institut d'Estadística de Catalunya

Please send this subscription form (or a photocopy) to:

SORT *(Statistics and Operations Research Transactions)*
Institut d'Estadística de Catalunya (Idescat)
Via Laietana, 58
08003 Barcelona
SPAIN
Fax: +34-93-412 31 45

Bank copy

Authorisation for automatic bank withdrawal in payment for
SORT *(Statistics and Operations Research Transactions)*

The undersigned ______________________________

authorises Bank/Financial institution ______________________________

located at (Street Address) ______________________________

Zip/postal code ______________________ City ______________________

Country ______________________________

to draft the subscription to **SORT *(Statistics and Operations Research Transactions)*** from my account

number □□□□ □□□□ □□ □□□□□□□□□□

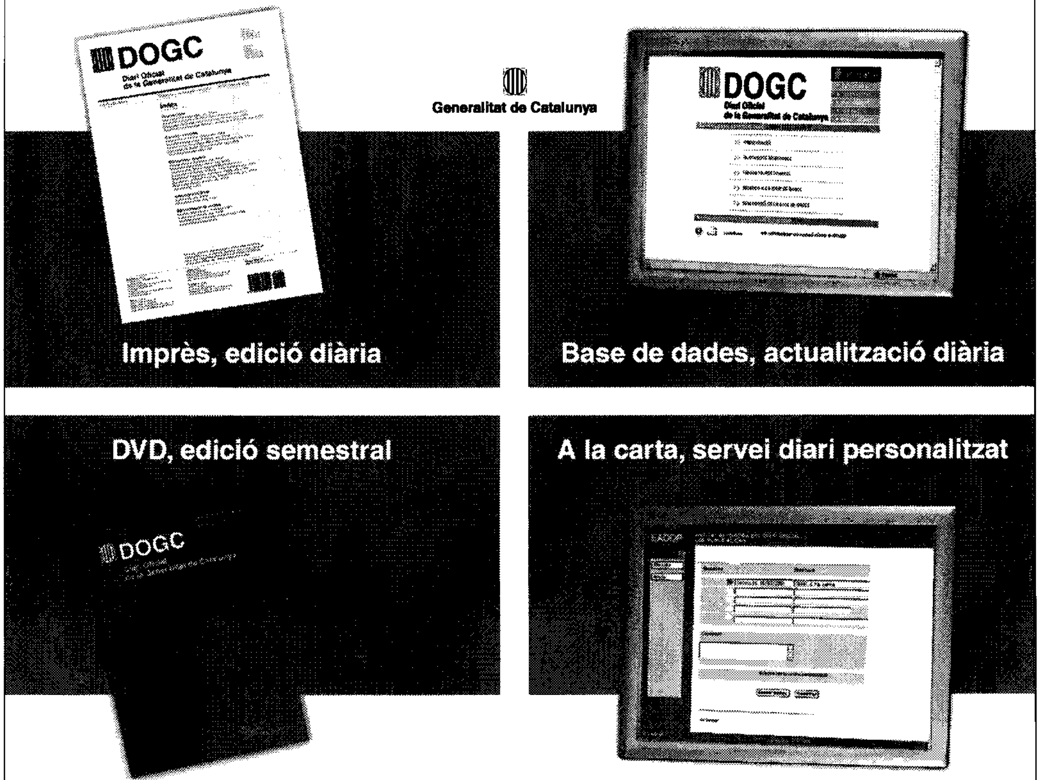
Date ______________________

Signature

SORT *(Statistics and Operations Research Transactions)*
Institut d´Estadística de Catalunya (Idescat)
Via Laietana, 58
08003 Barcelona
SPAIN
Fax: +34-93-412 31 45

Quatre modalitats de subscripció al DOGC

(Diari Oficial de la Generalitat de Catalunya)



A més, per als subscriptors de l'edició impresa i del DVD, tramesa gratuïta d'un CD-ROM trimestral que conté les pàgines en format PDF (DOGC en imatges)

L'Administració més a prop

EADOP • Informació i subscripcions • Rocafort, 120 - Calàbria, 147 • 08015 Barcelona
Tel. 93.292.54.17 • Fax 93.292.54.18 • subsdogc@gencat.net • www.gencat.net/eadop

Institut d'Estadística de Catalunya

The Institut d'Estadística de Catalunya (Idescat) is the statistical office of the Government of Catalonia. Its main duty is the management of the Catalan Statistical System by planning, coordinating and standardizing the statistical activity as well as providing statistical technical assistance.

Secretary and subscriptions to SORT

Institut d'Estadística de Catalunya (Idescat)
Via Laietana, 58
08003 Barcelona (Spain)
Tel. +34 - 93 412 09 24 - 93 412 00 88
Fax. +34 - 93 412 31 45
E-mail: sort@idescat.es

Publisher: Institut d'Estadística de Catalunya (Idescat)
© Institut d'Estadística de Catalunya
ISSN 1696-2281
DL B-46.085-1977
Key title: SORT
Numbering: 1 (december 1977)
http://www.idescat.net/sort

SORT

Volume 28 (2), July - December 2004
Formerly Qüestiió

Contents

Articles

ISSN: 1696-2281
http://www.idescat.es/sort

28002

9 771696 228009